



The Evolutionary Feedback between Genetic Conflict and Genome Architecture

Citation

Young, Adrian. 2014. The Evolutionary Feedback between Genetic Conflict and Genome Architecture. Doctoral dissertation, Harvard University.

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:12274113>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

The Evolutionary Feedback between Genetic Conflict and Genome Architecture

A dissertation presented

by

Adrian Young

to

The Department of Organismic and Evolutionary Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biology

Harvard University

Cambridge, Massachusetts

May 2014

© 2014 *Adrian Young*
All rights reserved.

The Evolutionary Feedback between Genetic Conflict and Genome Architecture

Abstract

The advent of separate sexes set the stage for dramatic evolutionary innovation across a wide range of taxa. Much of this innovation is attributable to divergent evolutionary interests between now distinct sub-populations of males and females. Trade-offs inherent to these divergent life histories, coupled with a common genome, conspire to limit natural selection's ability to simultaneously maximize the fitness of both sexes. Such conflict between the sexes has therefore largely shaped the history of the genomes of sexual taxa. However, various aspects of the genomic environment—including genes' spatial distributions, abilities to regulate their expression, and rates of recombination—also feed back to influence future sex-specific evolutionary trajectories. Using various genomic resources and transcriptome sequences for the lab mouse, I test several theoretical predictions regarding this feedback between genetic conflict and features of genomic organization.

I begin first by testing recent predictions regarding the clustering of loci experiencing sexual antagonism. Using methods from spatial ecology, I show that autosomal, but not X-linked, sex-biased genes form numerous clusters of limited spatial extent. This spatial organization likely facilitates the invasion of new sexually antagonistic alleles, yet may counterintuitively accelerate sex-specific adaptation as well. Then, I test common intuition on the role of hormones in promoting sex-biased expression. Contrary to expectations, I find that the canonical model of hormone-mediated transcriptional control has little explanatory power, suggesting more of a role for *trans*-acting genes in providing genes direct information on their

sexual environment. These results also suggest that the resolution of sexual antagonism may proceed more slowly than one might expect from the rate at which sex-biased expression evolves. Finally, I also test recent predictions on the relationship between sex-biased recombination rates and genomic imprinting. Consistent with theory, I find imprinted genes are associated with greater sex differences in recombination while controlling for other molecular features of the genome. I then discuss how such associations can couple the future trajectories of genes experiencing sexual and parental antagonism despite the independent ecological origins of such selection pressures.

TABLE OF CONTENTS

CHAPTER 1

| | |
|--|---|
| <i>The snags of separate sexes: conflict and coevolution</i> | 1 |
|--|---|

CHAPTER 2

| | |
|---|---|
| <i>Spatial ecology of the genome: sexual antagonism and linkage</i> | 8 |
|---|---|

CHAPTER 3

| | |
|---|----|
| <i>Hormones, sexual dimorphism, and the resolution of sexual antagonism</i> | 53 |
|---|----|

CHAPTER 4

| | |
|--|----|
| <i>Heterochiasmy and genomic imprinting: evolutionary and molecular perspectives</i> | 77 |
|--|----|

Acknowledgments

My thanks go first and foremost to my advisor, David Haig. David has altruistically provided me a wonderful academic home these last several years, and during that time has had a tremendous influence on my view of biology and the nature of science. Upon joining the lab, I set out with the task of matching David's breadth of knowledge and depth of understanding. Those who have had the pleasure of working with David know that such a task is more suited to Sisyphus than to any student. All these years later, I am happy to report that I have indeed failed miserably in this regard, but have grown immensely in the process. If I can somehow retain a fraction of the insight David has offered over time, I have no doubt that my own students will ultimately benefit over the course of my future teaching career.

Many thanks also go to Edgar Duéñez-Guzmán, as without his friendship, guidance, and enthusiasm, this dissertation would look very different and likely would have been completed several years too late. For the questions I have worked on, I can think of no better colleague than someone who can credibly and simultaneously wear the hats of a mathematician, computer scientist, and evolutionary theorist while also exhibiting extraordinary patience in the face of my frequent questions and other random thoughts.

I have also been blessed with other phenomenal colleagues, officemates, and friends over the years across the MCZ. Specifically, Manus Patten, Wenfei Tong, Mark Cornwall, Jesse Weber, Evan Kingsley, Lukas Rieppel, Yoel Stuart, Collin Johnson, Terry Ord, Sarah Kocher, Lucas Mix, Gabe Gartner, and Jen Kotler have had a profound and unique influence on my life in graduate school. I am also grateful to Eileen Lacey for being such a positive influence on me during my formative undergraduate years. The research in this dissertation is in some ways the

culmination of some of my experiences during that time. Clearly, one should never underestimate what tuco-tucos, a scale, and some old field notes can do for a budding career.

I would also like to thank the department of Organismic and Evolutionary Biology for all the funding, support, and opportunities it has provided. Specifically, I would like to thank Chris Preheim and Pam Greene for all they have done in finding me teaching opportunities and generally making life smoother in OEB. Also, I would like to thank my committee members past and present—Bill Bossert, John Wakeley, Naomi Pierce, and Dan Hartl—for all the challenges, guidance, and support they have provided.

Finally, none of this would have happened without the love and unending support of my family. Part of my interest in parent-offspring conflict and sibling rivalry must come from having experienced essentially none of it over the years. Such an environment gave me the strength to survive many periods of turbulence and ultimately got me to where I am today. For this, I am eternally grateful, and I therefore dedicate this dissertation to them.

CHAPTER 1

The snags of separate sexes: conflict and coevolution

The evolution of sex is one of the oldest and largest problems within biology. Although much work over the decades has shed light on the variety of factors favoring the maintenance of sex, its origins have remained insufficiently explained (Michod and Levin 1988; Otto 2009). However, a recent synthesis of evidence and ideas represents a notable exception to this trend. A central underlying feature of this synthesis is a form of genetic conflict involving the related, yet independent process of recombination. Specifically, selfish genetic elements in a primitive diploid asexual organism could benefit themselves by inducing a recombination event in which they replace their homolog through the process of gene conversion. Such elements thereby eliminate the benefits of diploidy, with the resulting loss of complementation of recessive deleterious alleles imposing a cost on the rest of the genome. This in turn would favor the evolution of syngamy and outcrossing to restore complementation, incidentally facilitating the spread of the initial selfish element as well (Archetti 2010). In such a scenario, sexual reproduction would be immediately advantageous, and would remain so in the face of such spreading selfish elements. In this way, sex evolves in response to intragenomic conflict yet also influences the context of future conflict.

With the evolution of sex, the evolution of *sexes* soon followed. Specifically, with syngamy as a normal part of the life cycle, natural selection in larger multicellular organisms could favor divergent investment strategies in the size of gametes (Parker et al. 1972). The ultimate consequence of such divergent selection was division of the population into those producing either many, small and motile gametes or producing few, large and non-motile gametes—sub-populations otherwise known as males and females. As small as this initial asymmetry may have been, it set the stage for a dramatic divergence of evolutionary interests between the sexes (Trivers 1972). Although reproduction is generally a cooperative venture,

males initially were essentially parasites that simply took advantage of greater female gametic investment. What then ensued was a long cycle of coevolution between the sexes affecting an increasingly large set of phenotypes directly or indirectly related to the process of reproduction. Whether over the process of finding, judging, deceiving, subjugating or fleeing potential mates prior to mating, or over the extent of parental investment and partner fidelity after mating, selection can and has directly favored markedly different strategies between the sexes (Arnqvist and Rowe 2005).

Indirect consequences of this conflict between the sexes are also widespread. In particular, basic life-history theory posits that the overall optimization of fitness by natural selection is subject to trade-offs between survival and reproduction (Stearns 1992). Intense selection acting on the various reproductive phenotypes listed above will therefore necessarily have correlated consequences for other survival-related phenotypes within each sex, independent of the particular genomic architecture of these traits. Additionally, large positive intersexual genetic correlations resulting from the sexes' shared genome ensure that any evolution along these life-history axes in one sex feeds back on the other. However, details of the underlying genetics also matter. Specifically, sex-chromosomal influences and pleiotropic relationships among phenotypes, as well as epistasis and patterns of physical linkage between different genes all underlie these genetic correlations. These features shape modern evolutionary trajectories, yet their abilities to do so derive in part from their lineage-specific histories and previous selective environments. In this way, conflict both past and present casts a long shadow over much if not all of the genomes of sexual taxa (e.g. Connallon and Clark 2014).

To develop this notion further, it may first help to slightly expand and modify our analogy of the sexes as separate populations competing against one another over the terms of

survival and reproduction. Particularly, it is important to recognize that, no matter the magnitude or scope of conflict they experience, the sexes as populations freely exchange migrants every generation. Over the span of their residency in the population, most alleles rapidly move back and forth between male and female bodies, spending approximately half their time in each (Haig 2002). From a gene's perspective, each sex is therefore an environment of sorts. Most migrant genes, however, are blind to the nature of both their previous and current niche. Under normal circumstances, such ignorance may be of no selective consequence. But under the shadow of conflict, genes may often find their single, unconditional phenotype ill-suited to their current sexual environment. This may be true even if these genes are simply bystanders, genetically correlated to those actively participating in such conflict. In this way, the load such conflict imposes can then favor the evolution of awareness and strategy, allowing genes to exhibit conditional behavior as they move back and forth between antagonistic sexual environments.

The evolution of two sexes was not limited solely to the production of males and females, however. The advent of two sexes also resulted in the evolution of mothers and fathers. Such an explicit distinction may initially seem unnecessary as all mothers are females and all fathers are males. However, differences between the ecology of parenthood and sexuality motivate a separation of the evolutionary interests of fathers from those of males, for example. This separation is particularly relevant for situations dominated by social interactions among kin within and between generations (Haig 2000; Wilkins and Haig 2003; Úbeda and Gardner 2010; 2011). Within a population, parental promiscuity (among other potential factors) creates an asymmetry in matrilineal and patrilineal relatedness among kin. Specifically, in a litter of half-sibs, maternally-derived alleles exhibit high relatedness, while the genetic contributions of multiple fathers render sibs' paternally-derived alleles relative strangers to one another. As

described above, though, most genes remain ignorant of their parental origin. Alleles acting in a focal individual to bias parental investment towards itself at a cost to its kin would be favored when paternally derived yet selected against when maternally derived. And as alleles spend equal time coming from mothers as they do from fathers, this particular allele essentially plays the ‘wrong’ strategy half the time. In response to such parental antagonism, natural selection would again favor genes’ strategic awareness (Haig 1997). Specifically, selection would favor the evolution of genomic imprinting, in which the alleles of a gene are either expressed or silenced depending on their sexual environment in the previous generation. Notably, as we did not specify the sex of our focal individual, such genes may be entirely agnostic as to the identity of their *current* sexual environment. Independent value in recognizing either past or present sexual environments implies that parental antagonism is not simply sexual antagonism one generation removed, despite these phenomena’s shared reliance on the existence of separate sexes.

Though distinctions between the interests of parents and the sexes may result from independent facets of their ecology, a variety of factors conspires to convolute their effects on the genome. For example, an ecological difference between the sexes related to dispersal can feed back directly onto parental interests. Mammals and birds, for instance, generally exhibit male and female-biased dispersal respectively. The ecology underlying mating system variation among these taxa is one factor thought to influence the evolution of such dispersal patterns (Greenwood 1980). However, sex-biased dispersal then feeds back, creating asymmetric patterns of matrilineal and patrilineal relatedness—particularly for genes on the X or Z chromosome (Haig 2000). These asymmetries can then influence the evolution of social behavior among philopatric individuals, likely contributed to in part by genes evolving imprinted

gene expression. Local features of the recombinational environment also serve to entangle the independent actions of parental and sexual antagonism. Depending on the degree of recombination, stable linkage disequilibrium can result from the separate actions of each form of conflict on two linked genes without epistasis (Patten et al. 2013). Specifically, linkage produces associations between matched sex and sex-of-origin beneficial alleles (e.g. female and maternally beneficial alleles). This linkage effectively boosts the strength of selection acting on either gene despite the fact that such selective pressures originate from distinct ecological contexts.

From the preceding exposition, it is clear that the emergence of separate sexes set the stage for substantial evolutionary change and innovation along a number of biological dimensions, genomic or otherwise. Similarly, we see that the resulting changes to the architecture and organization of the genome can feed back on these sex-influenced selective pressures. Such feedback can modify the nature of selection itself as well as shape the means and tempo of future resolutions to such forms of genetic conflict. In the following chapters, I will explore this cyclical relationship further, integrating a variety of genomic data with recent theoretical predictions to assess both the magnitude and temporal nature of the interplay between sexual and genomic environments.

Acknowledgments

I would like to thank Jen Kotler for helpful commentary on a previous version of this chapter.

References

- Archetti, M. 2010. Complementation, genetic conflict, and the evolution of sex and recombination. *Journal of Heredity* 101: S21-S33.
- Arnqvist, G. and L. Rowe. 2005. *Sexual conflict*. Princeton University Press, Princeton, NJ.
- Connallon, T. and A. G. Clark. 2014. Evolutionary inevitability of sexual antagonism. *Proceedings of the Royal Society B: Biological Sciences* 281: 20132123.
- Greenwood, P. J. 1980. Mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour* 28: 1140-1162.
- Haig, D. 1997. Parental antagonism, relatedness asymmetries, and genomic imprinting. *Proceedings of the Royal Society B: Biological Sciences* 264: 1657-1662.
- Haig, D. 2000. Genomic imprinting, sex-biased dispersal, and social behavior. *Annals of the New York Academy of Sciences* 907: 149-163.
- Haig, D. 2002. *Genomic imprinting and kinship*. Rutgers University Press, New Brunswick, NJ.
- Michod, R. E. and B. R. Levin. 1988. *The evolution of sex: an examination of current ideas*. Sinauer Associates, Sunderland, MA.
- Otto, S. P. 2009. The evolutionary enigma of sex. *The American Naturalist* 174: S1-S14.
- Parker, G. A., R. Baker and V. Smith. 1972. The origin and evolution of gamete dimorphism and the male-female phenomenon. *Journal of Theoretical Biology* 36: 529-553.
- Patten, M. M., F. Úbeda and D. Haig. 2013. Sexual and parental antagonism shape genomic architecture. *Proceedings of the Royal Society B: Biological Sciences* 280: 20131795.
- Stearns, S. C. 1992. *The evolution of life histories*. Oxford University Press, Oxford, U.K.
- Trivers, R. 1972. Parental investment and sexual selection, pp. 136-179 in *Sexual Selection and the Descent of Man 1871-1971*, edited by B. Cambell. Aldine Publishing Company, Chicago, IL.
- Úbeda, F. and A. Gardner. 2010. A model for genomic imprinting in the social brain: Juveniles. *Evolution* 64: 2587-2600.
- Úbeda, F. and A. Gardner. 2011. A model for genomic imprinting in the social brain: Adults. *Evolution* 65: 462-475.
- Wilkins, J. F. and D. Haig. 2003. Inbreeding, maternal care and genomic imprinting. *Journal of Theoretical Biology* 221: 559-564.

CHAPTER 2

Spatial ecology of the genome: sexual antagonism and linkage

Abstract

Recent theory has demonstrated that sexual antagonism—selection acting on an allele in opposing directions between the sexes—may be a potent and widespread force structuring the genomes of sexual taxa. As sexual antagonism at one locus can facilitate antagonism at a second closely linked locus, this process may result in the spatial clustering of sex-biased genes along chromosomes. Using estimates of sex-biased gene expression and existing genomic resources for the lab mouse, we tested this prediction while controlling for the potentially confounding effects of paralogy between nearby genes. Consistent with theoretical predictions, we find significant clustering of sex-biased genes at typically short recombinational scales across all autosomes but notably not on the X chromosome. These results shed light on the mode and tempo of the resolution of such genetic conflict and have implications for future sexual antagonism and sex-specific adaptation. Namely, we argue that genomes arranged to more easily maintain antagonistic alleles (with their associated fitness load) may paradoxically experience faster long-term phenotypic divergence between the sexes.

Introduction

Genome organization and evolution across all taxa is subject to a variety of influences. Consequently, many studies have found that genomes are non-randomly structured, particularly regarding the spatial arrangement of genes of various types (e.g. co-expressed, co-regulated, essential genes, etc.) (Hurst et al. 2004). One such process that has been invoked recently as a possible influence on genome organization is sexual antagonism.

Sexual antagonism, also known as intralocus sexual conflict, is the phenomenon in which an allele is favored in one sex but deleterious in the other (Bonduriansky and Chenoweth 2009; van Doorn 2009). Previous theory has typically focused on a single locus, but a pair of recent studies investigated the dynamics of two independent, but linked loci that segregate for sexually antagonistic variants (hereafter referred to as sexually antagonistic loci for simplicity) (Patten et al. 2010; Úbeda et al. 2011). Two findings from their model are important here. One finding is that linkage disequilibrium can be stably maintained by a pair of antagonistic loci, particularly for lower recombination rates. Second, this linkage also contributes to fitness variation in excess of predictions from simpler single locus models. Therefore, new sexually antagonistic alleles are more likely to invade and remain polymorphic when closely linked to a previously existing antagonistic locus.

Stable polymorphism at these loci prevents either sex from reaching their respective fitness optima, creating the opportunity for selection to favor some resolution of this conflict. For a gene experiencing antagonism over expression levels, natural selection would thus favor the evolution of sex-biased expression as a resolution. Hence, a reasonable assumption is that genes with sex-biased expression represent loci experiencing ongoing or recently resolved sexual antagonism (Bonduriansky and Chenoweth 2009). This assumption, coupled with the theoretical

results described above, therefore presents us with the testable prediction that sex-biased genes should be spatially clustered in actual genomes.

Other processes besides sexual antagonism can also give rise to close spatial associations between genes of a certain kind. Namely, tandem duplication of sex-biased loci via unequal crossing over could be responsible for this predicted pattern. Studies have shown that gene duplication is one means by which selection can favor the resolution of sexual antagonism (Gallach et al. 2010; Connallon and Clark 2011). And by virtue of proximity to its parent, a tandem duplicate gene would also likely experience the benefits of close linkage detailed above. Therefore, any assessment of the spatial clustering of sex-biased genes should also take their potentially paralogous relationships into account.

Different modes of inheritance and their effects on linkage may also influence observable clustering patterns. Specifically, the prediction of clustering arises from a model of autosomal loci, but this prediction may not generalize to X-linked sex-biased genes. The X chromosome is both predicted under certain circumstances and known in certain species to be a hot-spot for sexual antagonism, and thus, sex-biased genes (Rice 1984; Gibson et al. 2002; Ellegren and Parsch 2007; Patten and Haig 2009; Connallon and Clark 2010; Fry 2010). However, male hemizyosity and female-limited recombination on the X may influence the production of any spatial patterns based on linkage among these genes.

Finally, it is worth considering what exactly is meant by clustering both here and in previous studies. Most studies of gene clustering with a distinct molecular perspective offer some sort of explicit definition of a spatial cluster. These definitions may be based either on the presence of runs of adjacent genes of the same type or on arbitrarily chosen distance thresholds measured in base pairs or numbers of genes (Lercher and Hurst 2006; Semon and Duret 2006;

Mank et al. 2008; Weber and Hurst 2011). While such definitions may have merit from a molecular perspective, their utility breaks down in the context of inheritance and selection. Instead, a notion of distance based on recombination (i.e. genetic map distances) would be more meaningful from an evolutionary perspective. Such a perspective may also suggest against forming any explicit definition or demarcation of a cluster, even one based on a proper distance metric. For instance, in the context of adaptation by natural selection, a gene has been defined as any segment of DNA that persists mostly unmodified such that it could serve as a unit of selection (Williams 1966; Dawkins 1976; Haig 2012). Given this definition's amorphous nature, it would seem unwise to ascribe arbitrarily concrete boundaries to a higher-order collection of such genes. Thankfully, the *behavior* of clustering is known as another concept that is not dependent upon distinct definitions of a cluster. Specifically, genes clustered together exhibit positive spatial autocorrelation, in which genes of a given type appear closer together than expected by chance.

Therefore, we intend to test the prediction that sex-biased genes should be positively spatially autocorrelated along their chromosomes, specifically using analytical approaches borrowed from spatial ecology. These approaches allow us to quantify these patterns from a genome-wide perspective and investigate the degree to which clustering behavior may decay with distance. Understanding these patterns will shed light on sexual antagonism's influence on the organization of the genome, which in turn may affect the evolution of future conflict and the process of sex-specific divergence.

Methods

Focal organism and genomic resources

To address these theoretical predictions, we made use of genomic resources available for the laboratory mouse. First, we utilized a recent, high-resolution linkage map constructed using over 10,000 SNPs and over 3,500 meioses from a large population of heterogeneous stock mice (Cox et al. 2009). This information was used to assign map positions to genes with coordinates in the UCSC Genome Browser mm9 genome (Fujita et al. 2011). Due to high marker density, we simply performed nearest-neighbor interpolation to assign these positions to genes located between the outermost marker SNPs on each chromosome (approximately 98% of autosomal and X-linked genes). We then gathered information regarding potential paralogy between these genes. Specifically, we used BioMart to download paralogous relationships estimated by Ensembl for roughly 56% of the genes investigated here (release 67; Vilella et al. 2009).

We also used data from a recent, large-scale study of the transcriptome in two separate regions of the adult *Mus* brain to estimate sex-biased gene expression (Gregg et al. 2010). Although this data was originally used in the context of genomic imprinting, the analyses described below are entirely novel and distinct from previous uses of the data. Illumina RNA sequencing reads for males and females were downloaded from the Sequence Read Archive and aligned with TopHat v. 1.3 to the mm9 genome (Trapnell et al. 2009; Kodama et al. 2012). Expression levels were thus ultimately characterized by counts of reads aligned to any transcript of a particular gene model in the UCSC mm9 knownGene table.

Sex-biased expression, for each gene in both brain regions, was evaluated by Fisher's exact test as described in Table 2.1 and elsewhere (Auer and Doerge 2010).

Table 2.1. Estimating sex-biased expression. Letters inside the table represent cell counts of aligned reads. The LOD statistic used to numerically represent expression bias is equal to $\log_e(AD/BC)$.

| | Male | Female |
|------------------------|----------|----------|
| Gene X | <i>A</i> | <i>B</i> |
| Remaining genes | <i>C</i> | <i>D</i> |

The natural log of the odds ratio (LOD) of this test's 2×2 contingency table was used as a measure of bias, with positive and negative LOD scores respectively indicating male and female-biased expression. The resulting distribution of p -values was used to calculate q -values to address the false discovery rate (Storey and Tibshirani 2003). Very small sex-biases at highly-expressed loci can be statistically significant even though the biases are unlikely to be biologically significant. Therefore, genes were only considered sex-biased if $q < 0.01$ and $|\text{LOD}| > \log_e(1.2)$. This LOD threshold was chosen based on a prior microarray study of whole-brain gene expression also in *Mus* (Yang et al. 2006). This particular study found that very few genes in the brain exhibited a 2-fold or greater sex-bias in expression. We reach a similar conclusion here using far more spatially precise RNA sequencing data, indicated that our more subtle threshold of expression should be appropriate in mediating the trade-off between false positives and negatives.

Using these approaches, we estimated that among the genes assigned genetic map positions here, 3,313 exhibited male-biased expression, 3,994 exhibited female-biased expression, and 206 genes were biased in opposing directions (male and female biases) across the two brain regions sampled, for an overall total of 7,513 sex-biased genes (Table 2.2). For subsequent analyses, simple binary values were assigned to all genes investigated here. Specifically, sex-biased genes were assigned a value of 1, and all other genes assigned a value of 0. These values, along with their associated fixed chromosomal locations, represent the data used to calculate spatial autocorrelations.

Table 2.2. A breakdown of the data on counts of sex-biased genes, recombinational distances, and basic gene content by chromosome based on previously published analyses.

| Chromosome | Male-biased | Female-biased | Doubly biased | Total genes | Span (cM) |
|-------------------|--------------------|----------------------|----------------------|--------------------|------------------|
| 1 | 261 | 179 | 11 | 1435 | 96.81 |
| 2 | 267 | 321 | 18 | 2566 | 102.00 |
| 3 | 204 | 265 | 7 | 1191 | 79.93 |
| 4 | 174 | 265 | 12 | 1701 | 86.05 |
| 5 | 189 | 271 | 14 | 1657 | 86.91 |
| 6 | 150 | 181 | 14 | 1592 | 76.95 |
| 7 | 158 | 416 | 15 | 2409 | 81.45 |
| 8 | 133 | 251 | 7 | 1283 | 74.18 |
| 9 | 201 | 204 | 11 | 1665 | 72.59 |
| 10 | 165 | 196 | 8 | 1383 | 75.50 |
| 11 | 180 | 400 | 16 | 1879 | 85.05 |
| 12 | 145 | 114 | 9 | 981 | 61.49 |
| 13 | 158 | 99 | 5 | 1025 | 64.20 |
| 14 | 130 | 114 | 7 | 1166 | 61.65 |
| 15 | 119 | 188 | 7 | 1098 | 57.03 |
| 16 | 132 | 106 | 4 | 756 | 54.66 |
| 17 | 109 | 255 | 13 | 1535 | 58.76 |
| 18 | 123 | 67 | 9 | 728 | 56.55 |
| 19 | 90 | 153 | 7 | 779 | 52.88 |
| X | 225 | 66 | 12 | 958 | 74.43 |

Measure of spatial autocorrelation

In order to evaluate the clustering behavior of sex-biased genes across a chromosome, we calculated a generalized version of Moran's I statistic (Moran 1948; Moran 1950; Cliff and Ord 1973). This statistic is a global measure of autocorrelation among values of a single variable existing at fixed positions in space. For our purposes, the variable of interest X is the presence or absence of sex-biased expression, with each x_i encoded as a 1 or 0. These binary values are then distributed across genes' fixed genetic map locations along each chromosome. Therefore, a chromosome simulated under the null hypothesis of spatial randomness would hold constant genes' locations, permuting their X values instead.

Although typically used for two-dimensional landscapes, Moran's I can be readily applied to the one-dimensional chromosomal environments examined here. The key consideration involves the construction of the generalized weight matrix, \mathbf{W} . The simplest typical version of a weight matrix is one in which $w_{ij} = 1$ if genes i and j (where $i \neq j$) are spatially adjacent, and 0 otherwise. However, as the spatial nature of inheritance depends on the recombination rate between loci, a proper weight matrix in this context should consider a notion of distance beyond simple adjacency. Additionally, a generalized matrix can consider more than just spatial relationships between genes. Therefore, we use a matrix that jointly considers both recombinational distances and paralogy between genes.

To represent spatial relationships, we defined the parameter d_{ij} to represent the recombinational distance between two genes. This value is calculated using the Kosambi mapping function (Kosambi 1943) to convert the centiMorgan (cM) distance between two genes into their associated recombination fraction. We then define the binary weight parameter g_{ij} to represent the spatial contribution to the overall weight matrix. This parameter takes the value of 1 when $d_{ij} \leq r_t$, where r_t is a threshold recombination fraction, and 0 otherwise. Similarly, we defined a binary weight parameter p_{ij} to represent the contribution of paralogy to the weight matrix. In this case, $p_{ij} = 0$ when genes i and j are each other's paralogs, and $p_{ij} = 1$ otherwise. Therefore, the w_{ij} values of our overall weight matrix can be calculated by simply taking the product $g_{ij} p_{ij}$ for all n genes on a chromosome. We also repeat analyses using a weight matrix that ignores paralogy in order to estimate its influence on overall spatial patterns. The resulting expression for Moran's I is given on the following page:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1, j \neq i}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\left(\sum_{i=1}^n \sum_{j=1, j \neq i}^n w_{ij} \right) \left(\sum_{i=1}^n (x_i - \bar{x})^2 \right)}$$

As I is asymptotically normally distributed with closed form expressions for its mean and variance, we determined significance by comparing its Z-score against a standard normal distribution. Although I superficially resembles a Pearson correlation coefficient, its lower and upper bounds depend instead on the structure of the data and the weight matrix used. Therefore, to assess the effect size of autocorrelation that would be comparable across chromosomes, we calculated a standardized form of I that falls within ± 1 as specified by (Upton and Fingleton 1985). Similar calculations were also performed for the subset of the data representing only sex-biased genes. In this case, male-biased genes received the value of 1 and female-biased genes the value of 0, while remaining genes' data and map positions were omitted from the analysis. In contrast to the more general analysis, these calculations address potential spatial associations *within* the category of sex-biased genes.

Spatial extent of autocorrelation

The construction of the spatial weight matrix described above is critically dependent on the choice of r_t , the threshold recombination fraction. Theoretical results suggest that the forces promoting clustering do so most strongly at short scales with their effects declining precipitously as recombination increases (Patten et al. 2010; Úbeda et al. 2011). However, no natural choice of a single threshold emerges from these results, nor is there an obvious sense of what an upper limit on 'short' recombinational distances might be. Therefore, we investigate potential

clustering across a range of reasonable recombination thresholds, in a manner somewhat analogous to the use of correlograms in traditional spatial statistics.

Correlograms allow one to investigate autocorrelation at a variety of scales in which long-range patterns (e.g. periodicity) are separated out from those occurring at shorter spatial scales. This separation requires data to be arranged into discrete distance classes in a meaningful way (i.e. adjacent neighbor, second most adjacent neighbor, etc.) (Cliff et al. 1975). However, the continuous nature of recombinational distances between genes would preclude such an approach, and a biological interpretation of distinct long-range autocorrelation may be of questionable value here.

Therefore, we instead construct ‘pseudo-correlograms’ to demonstrate how patterns of spatial autocorrelation change as the threshold recombination fraction increases. In other words, we investigate short-range spatial autocorrelation using an increasingly liberal definition of ‘short.’ Specifically, we make our calculations with 20 values of r_t that range from 0 to 0.19 in increments of 0.01. We chose the maximum value for this range based on the observed genome-wide average pair-wise recombination fraction. Although this is an arguably arbitrary metric to rely on for our choice, this upper boundary is worthwhile for several reasons. First, this value allows for an examination of spatial patterns across a fairly wide range of recombinational distances. Second, at this scale, genes within a 22 cM radius of each other would be considered close. Considering that an average *Mus* chromosome spans about 73 cM, the notion of close proximity has already lost considerable meaning for this upper boundary of recombination, suggesting against the use of even higher values. Finally, this particular boundary is useful in limiting the degree of multiple testing, as the same underlying data is being used for every value of r_t .

Specifically, 20 tests are performed across this range of r_i for both forms of the weight matrix (constructed with and without considering paralogy), giving a total of 40 tests per chromosome. In order to prevent this from inflating our overall Type I error rate, we use a Bonferroni-corrected α of 0.00125 (i.e. 0.05/40) as our threshold of significance. For our main test of sex-biased gene clustering, we make use of a one-tailed significance threshold as we are explicitly testing the alternate hypothesis of *positive* spatial autocorrelation. For the standard normal distribution used here, this threshold corresponds to a value of Z approximately equal to 3.023. In contrast, we have no prior basis for a one-tailed test of spatial autocorrelation *within* the subset of sex-biased genes. Therefore, we make use of a two-tailed test, corresponding to Z thresholds of ± 3.227 . Finally, all data processing and statistical testing was performed using custom Python scripts.

Results

Autocorrelation among sex-biased genes

In agreement with theoretical predictions, we find significant but modest positive spatial autocorrelation among sex-biased genes across the entire genome (Figure 2.1). This result is also largely robust to the potentially confounding effects of paralogy. Excluding gene pairs whose spatial relationship resulted from either tandem duplication or transposition reduced the magnitude of autocorrelation slightly, but the overall pattern remained intact across the entire range of recombination thresholds. Considering these thresholds, we found maximal spatial autocorrelation at the most restrictive recombinational scales, typically that of complete linkage with $r_i = 0$ (Table 2.3).

Table 2.3. Table of maximum standardized Moran's I values and their associated recombination thresholds (r_t) for the general spatial autocorrelation analysis. As described in the text, $I \in [-1, +1]$ and $r_t \in [0, 0.19]$ and the weight matrix used is given in parentheses. Positive values of I indicate clustering of like genes (e.g. biased genes with biased genes).

| Chromosome | I (distance) | r_t (distance) | I (distance + paralogy) | r_t (distance + paralogy) |
|--------------------------|----------------|------------------|---------------------------|-----------------------------|
| 1 | 0.144 | 0 | 0.133 | 0.01 |
| 2 | 0.251 | 0 | 0.239 | 0 |
| 3 | 0.146 | 0 | 0.131 | 0 |
| 4 | 0.258 | 0 | 0.241 | 0 |
| 5 | 0.229 | 0 | 0.222 | 0 |
| 6 | 0.209 | 0 | 0.199 | 0 |
| 7 | 0.302 | 0.01 | 0.272 | 0.01 |
| 8 | 0.222 | 0 | 0.213 | 0 |
| 9 | 0.219 | 0 | 0.201 | 0 |
| 10 | 0.245 | 0 | 0.240 | 0 |
| 11 | 0.165 | 0 | 0.131 | 0 |
| 12 | 0.247 | 0 | 0.245 | 0 |
| 13 | 0.224 | 0.02 | 0.208 | 0.02 |
| 14 | 0.236 | 0.01 | 0.189 | 0.01 |
| 15 | 0.255 | 0 | 0.252 | 0 |
| 16 | 0.180 | 0 | 0.121 | 0 |
| 17 | 0.295 | 0.01 | 0.287 | 0.01 |
| 18 | 0.266 | 0 | 0.265 | 0 |
| 19 | 0.277 | 0.01 | 0.264 | 0.01 |
| X | 0.088 | 0.19 | 0.069 | 0.19 |
| Autosomal average | 0.230 | 0.003 | 0.213 | 0.004 |

At increasingly liberal distances, autocorrelation tended to decay below our significance threshold fairly quickly for a majority of chromosomes (Figure 2.1).

The X chromosome stands in stark contrast to the general patterns that we observe among the autosomes. Although we do find positive spatial autocorrelation on the X, this signal does not decay with distance and in fact only emerges for larger values of r_t . Notably, the X chromosome attains its maximum value of autocorrelation at our uppermost recombination

boundary. However, this maximum value is only about one-third that observed on the autosomes (Table 2.3).

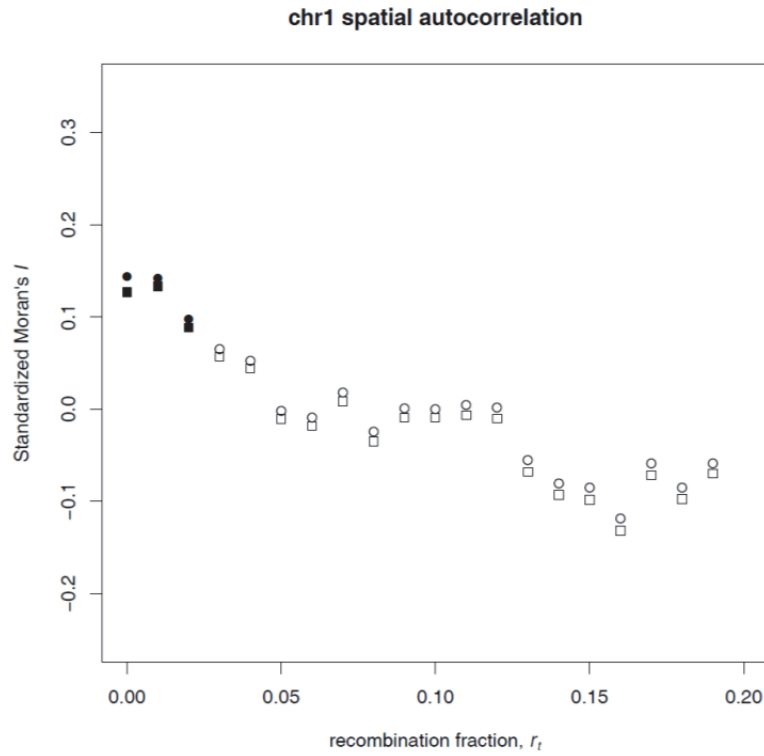


Figure 2.1. 'Pseudo-correlograms' representing the clustering of sex-biased genes across all values of the threshold recombination fraction, r_t , for all chromosomes. Circles represent standardized autocorrelation values for the distance-only analysis and squares represent values for the analysis considering both distance and paralogy. Filled symbols indicate values significantly greater than zero, with the Z-score of the point of maximal autocorrelation given explicitly within the graph.

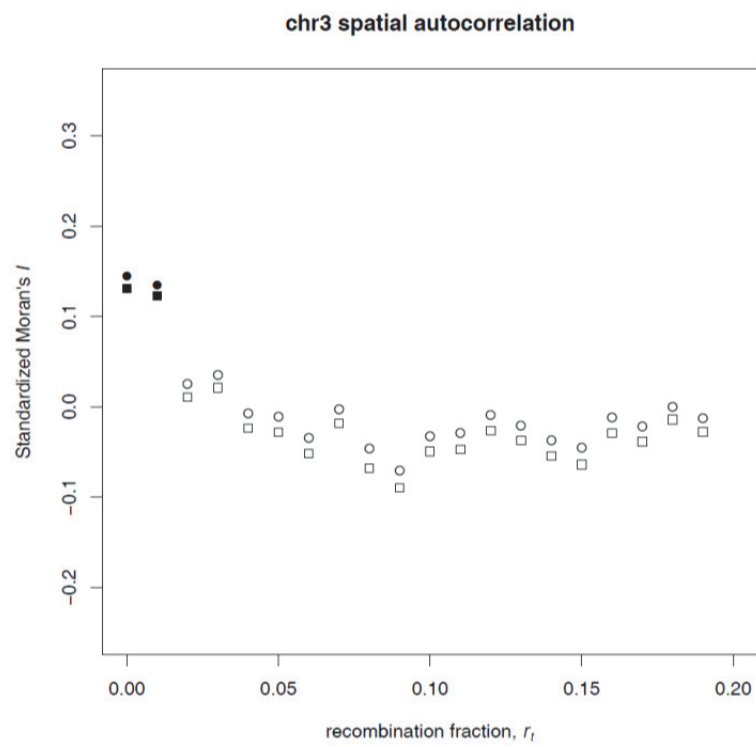
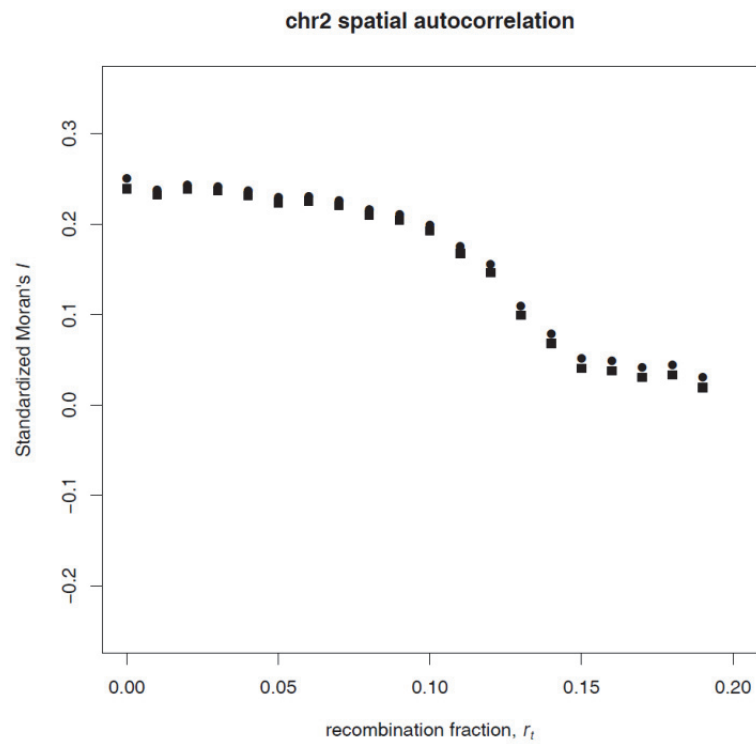


Figure 2.1 (Continued)

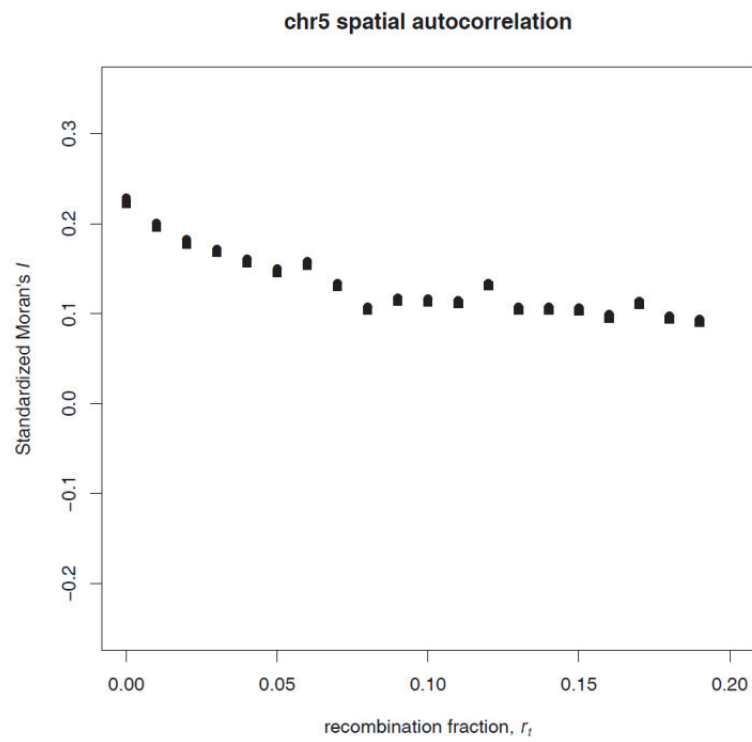
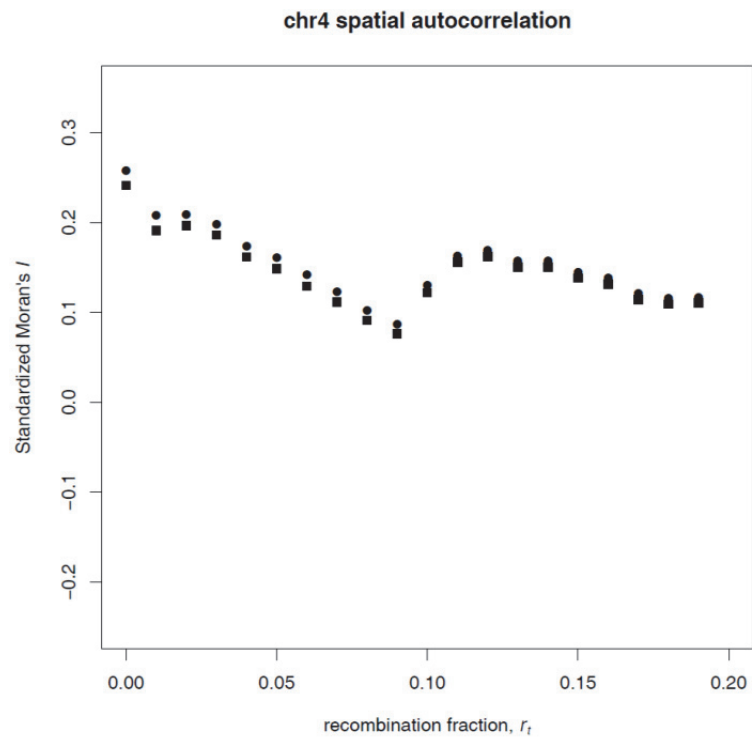


Figure 2.1 (Continued)

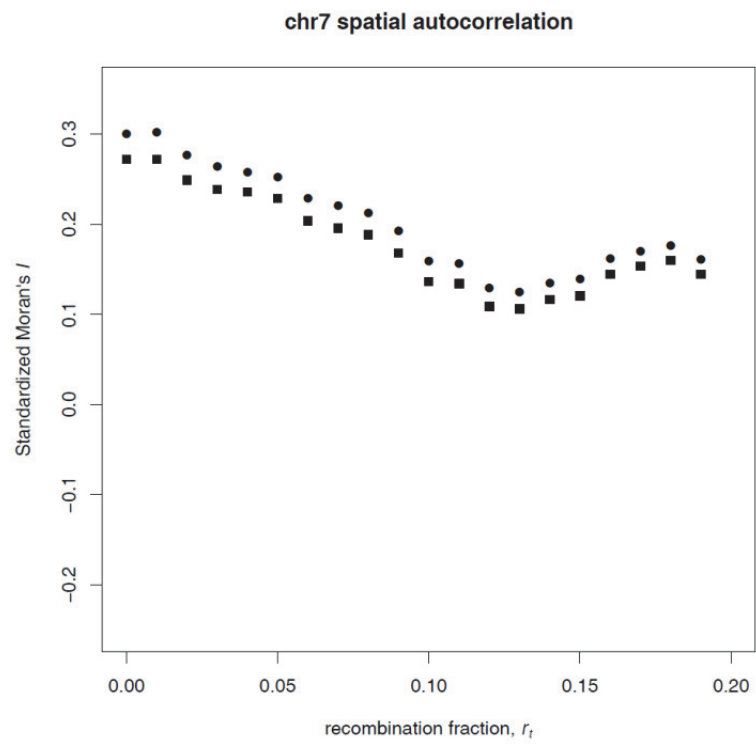
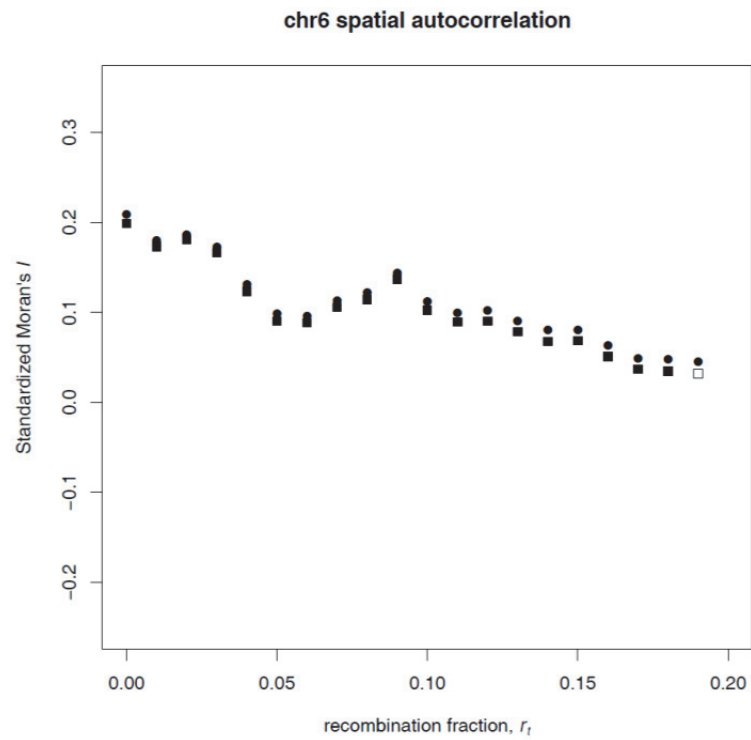


Figure 2.1 (Continued)

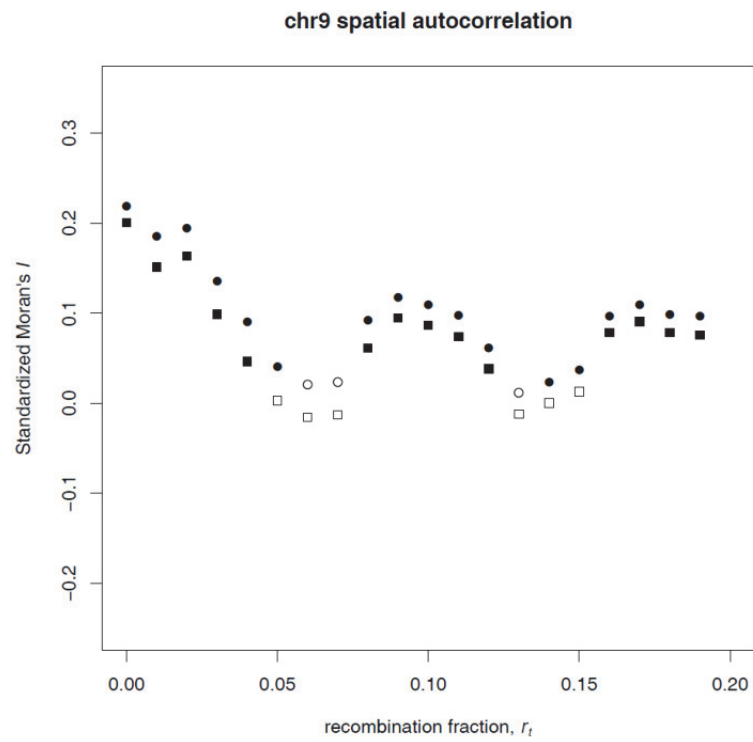
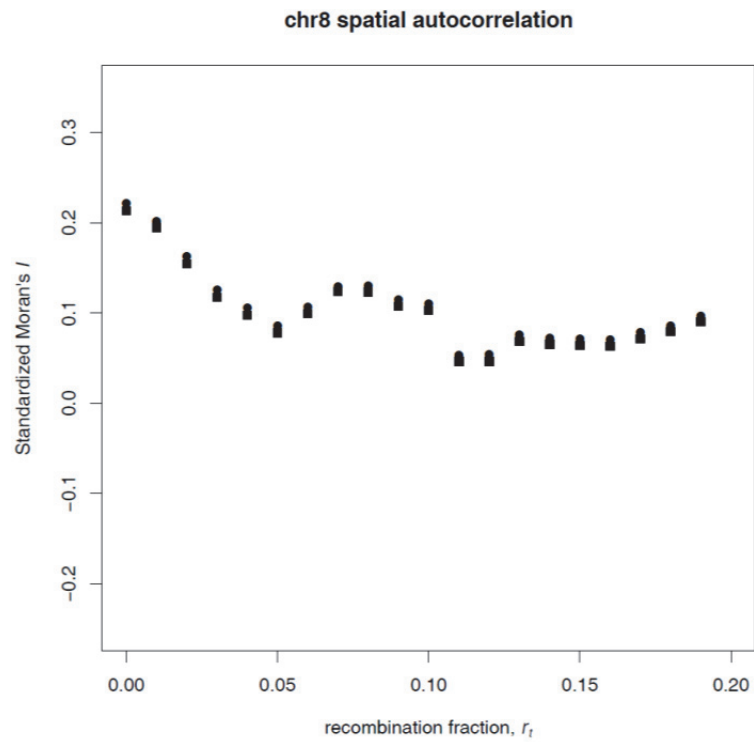


Figure 2.1 (Continued)

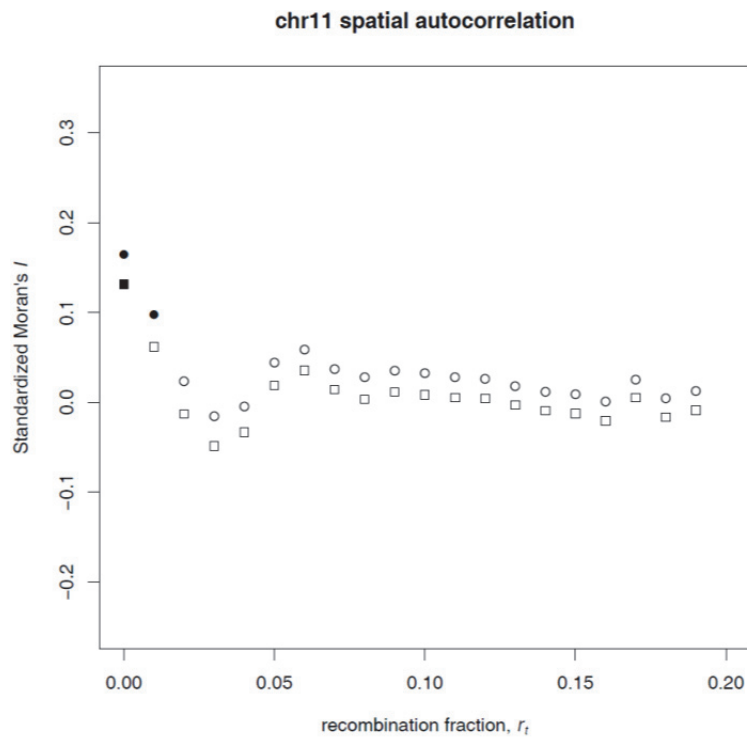
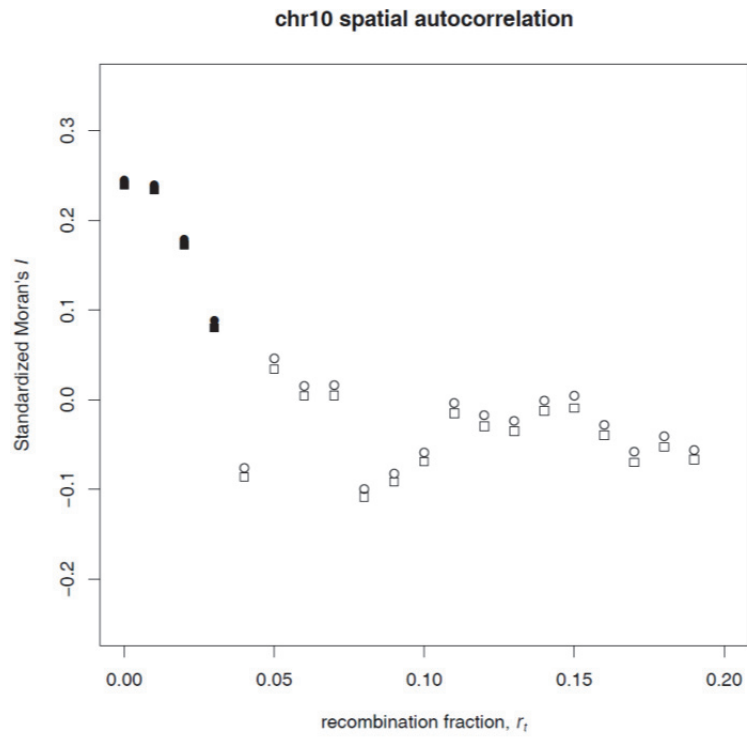


Figure 2.1 (Continued)

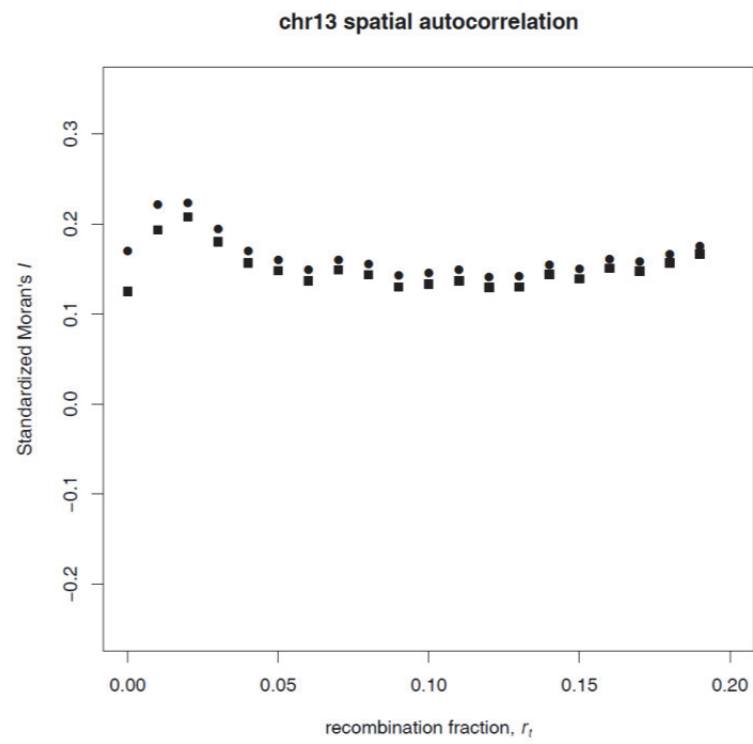
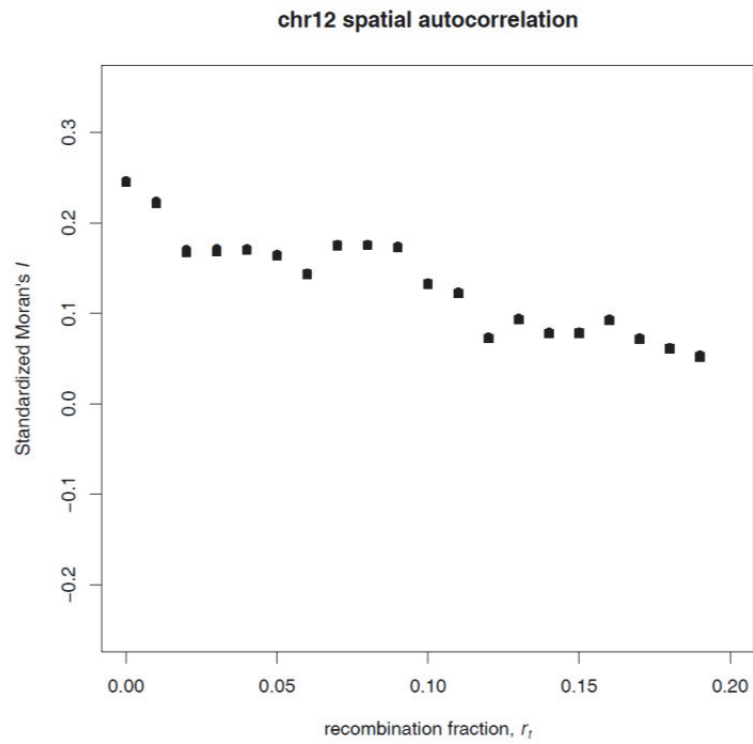


Figure 2.1 (Continued)

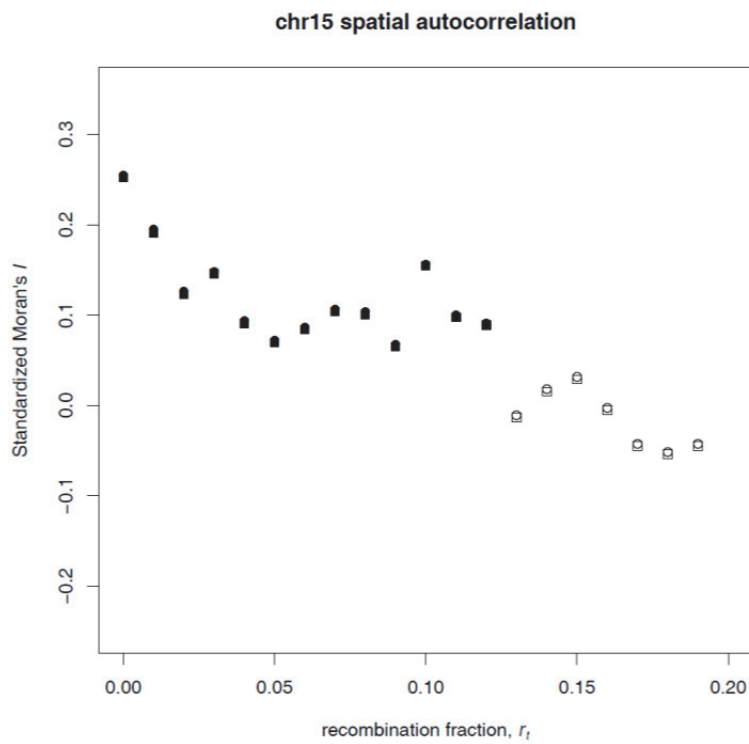
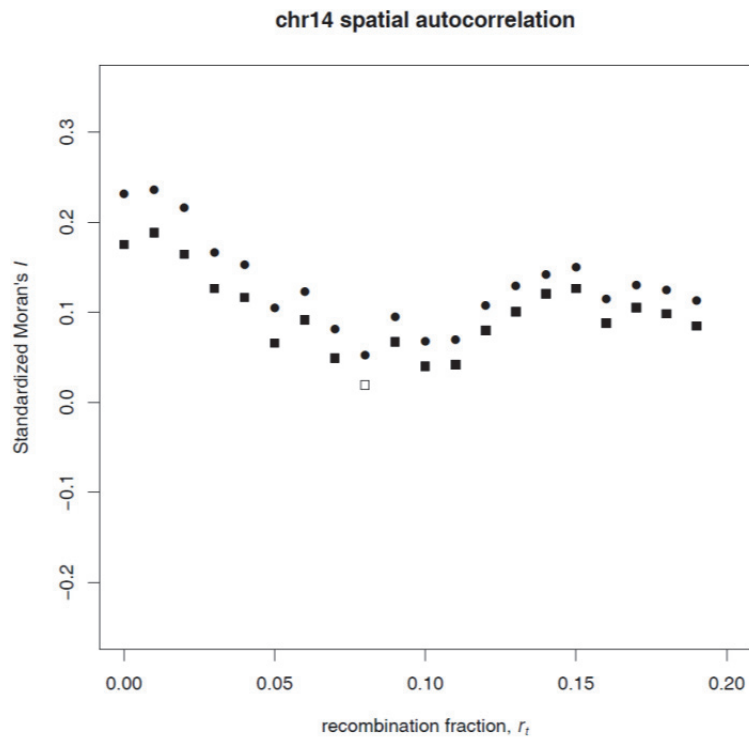


Figure 2.1 (Continued)

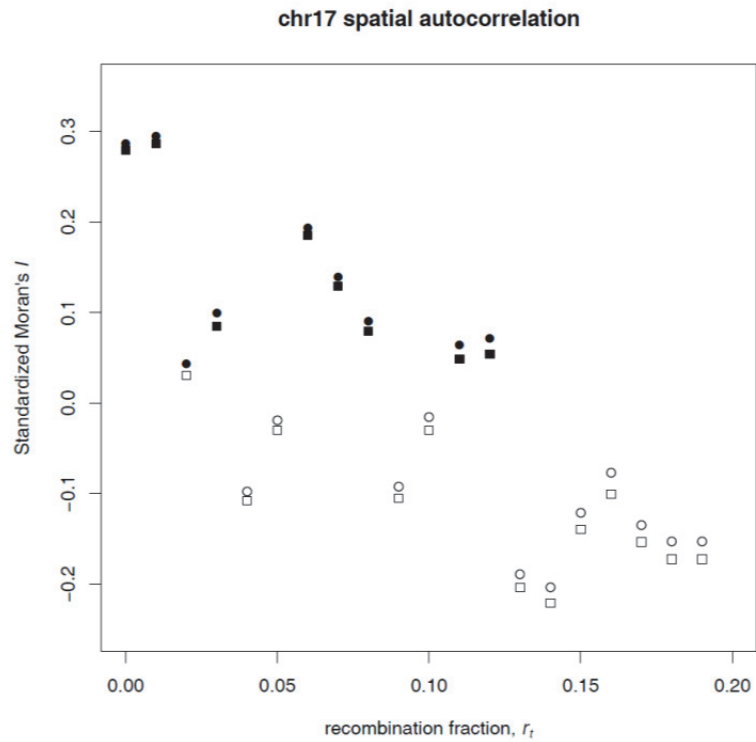
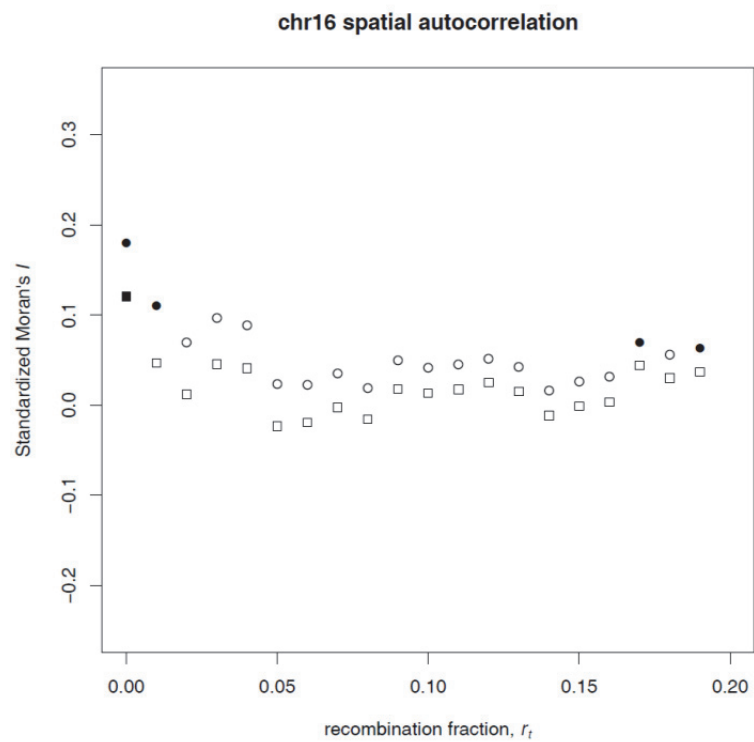


Figure 2.1 (Continued)

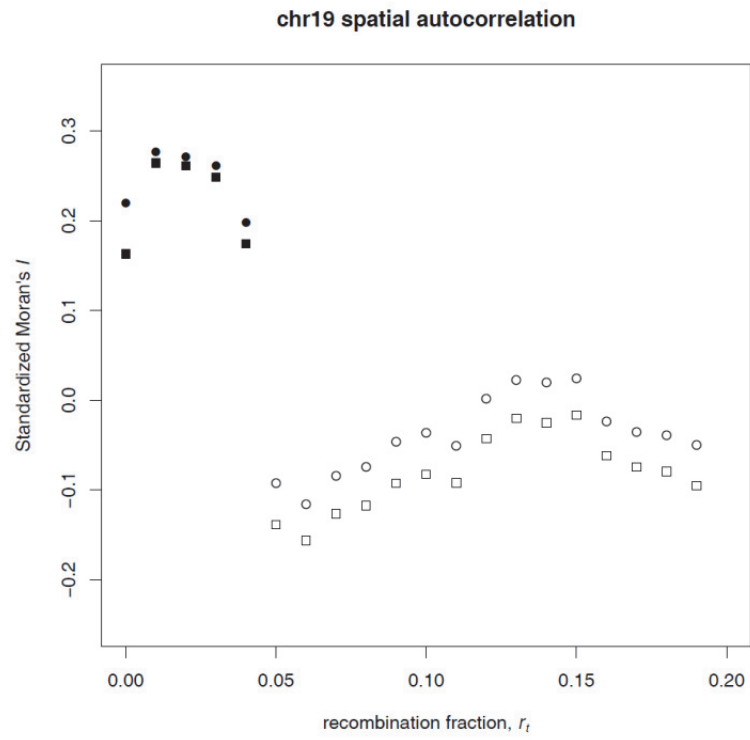
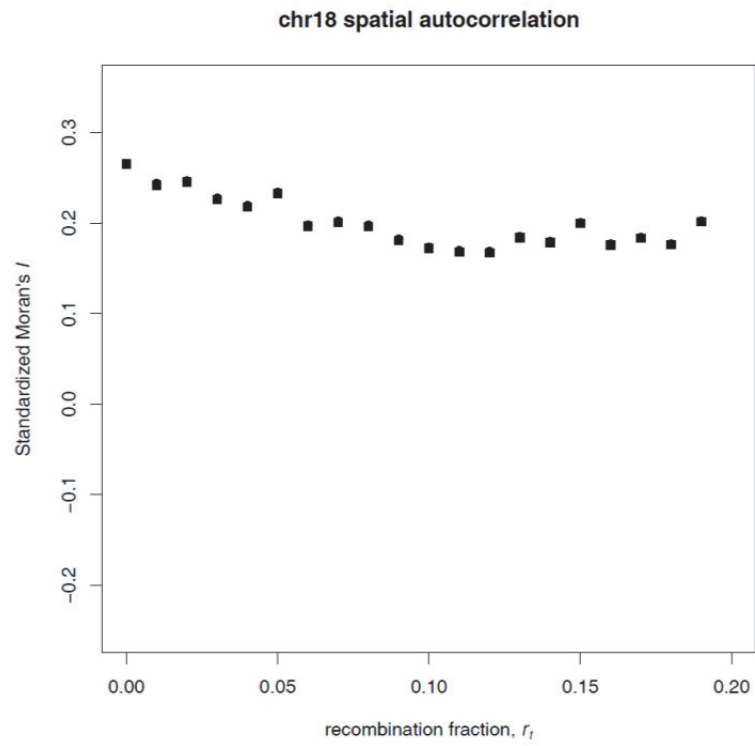


Figure 2.1 (Continued)

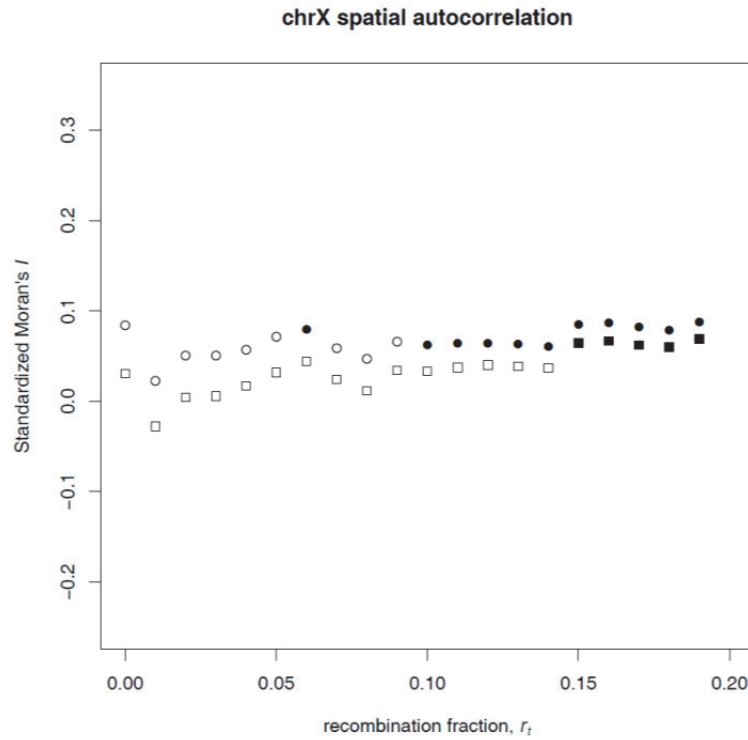


Figure 2.1 (Continued)

Autocorrelation within sex-biased genes

Roughly similar patterns are also evident within the category of sex-biased genes. Namely, if we restrict our attention to the spatial relationships between male and female-biased genes, we again find significant positive autocorrelation minimally affected by paralogy (Figure 2.2).

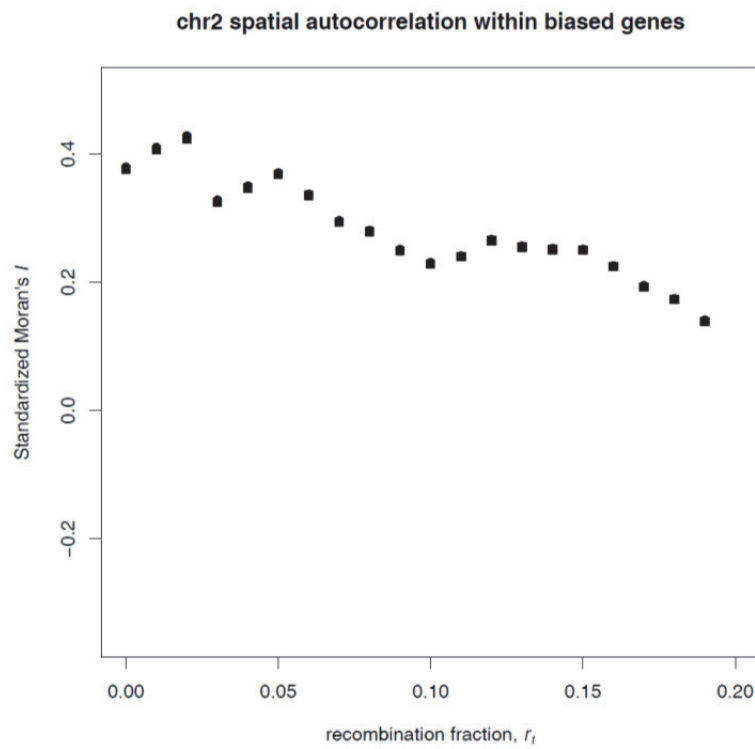
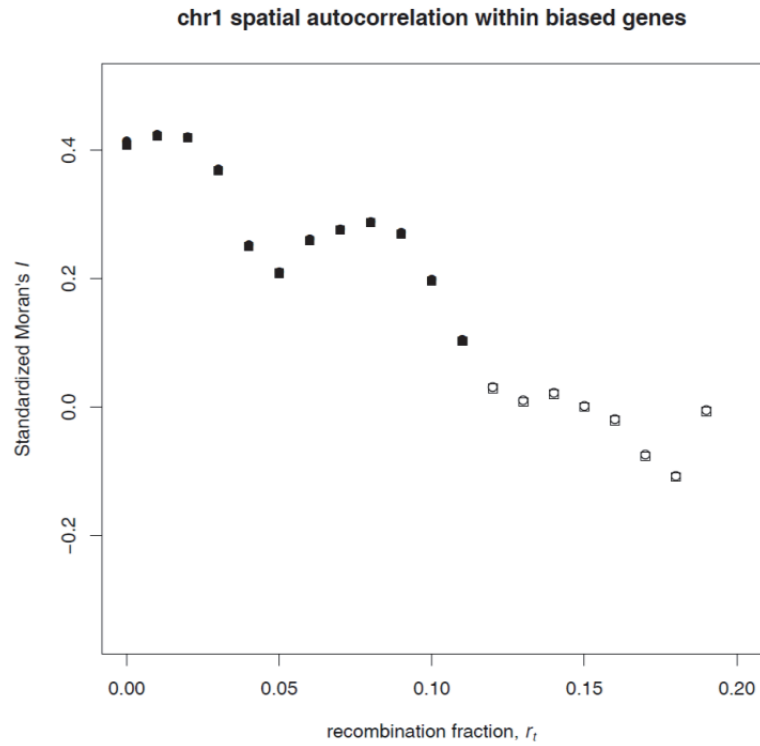


Figure 2.2. ‘Pseudo-correlograms’ representing the clustering *within* sex-biased genes for all chromosomes. Features of the graphs are identical to those described in Figure 2.1.

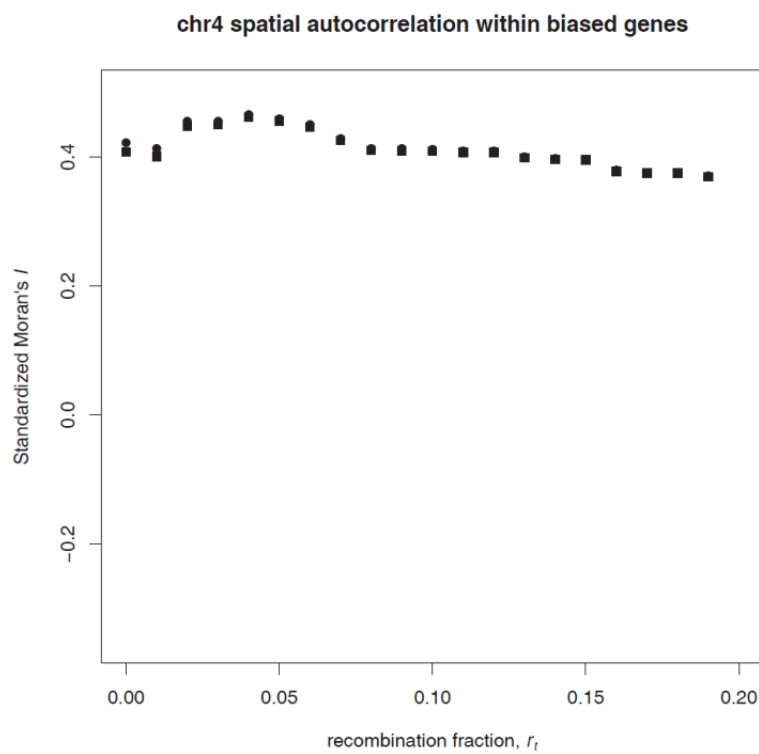
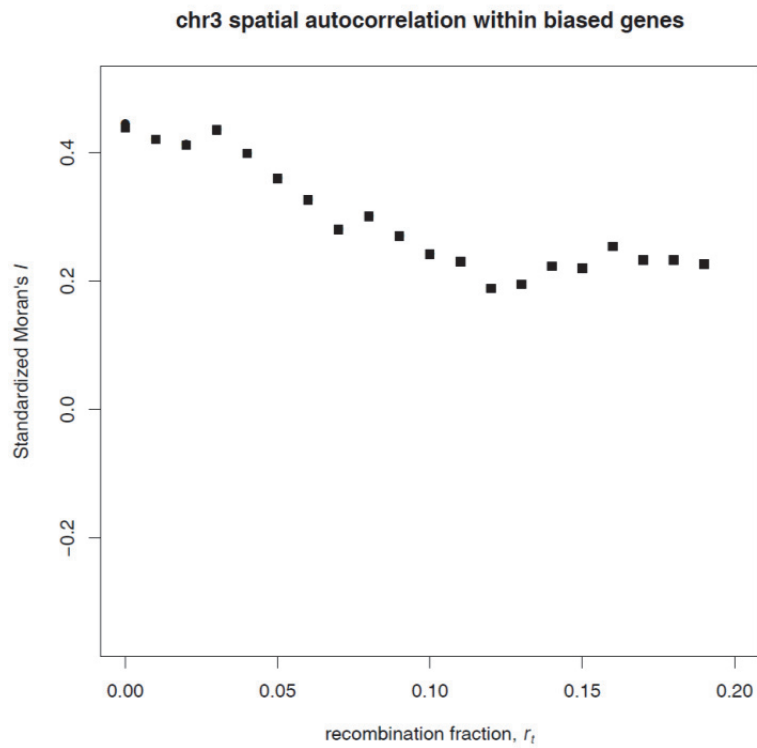


Figure 2.2 (Continued)

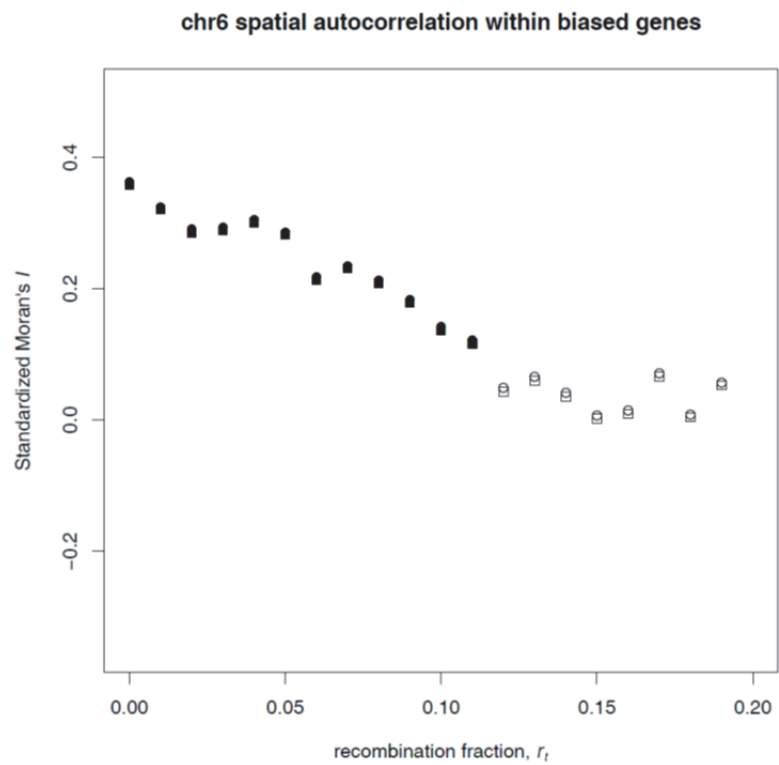
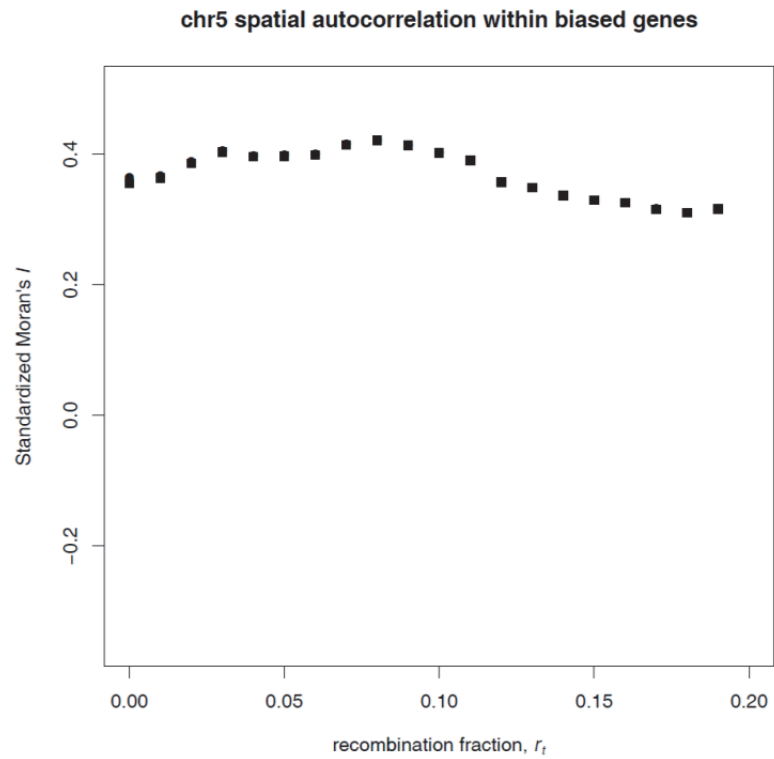


Figure 2.2 (Continued)

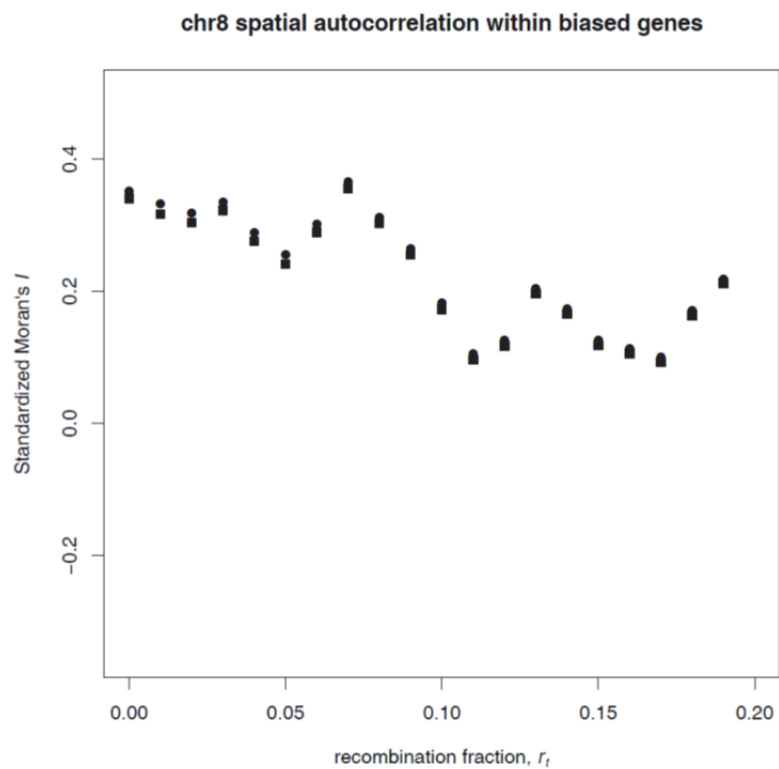
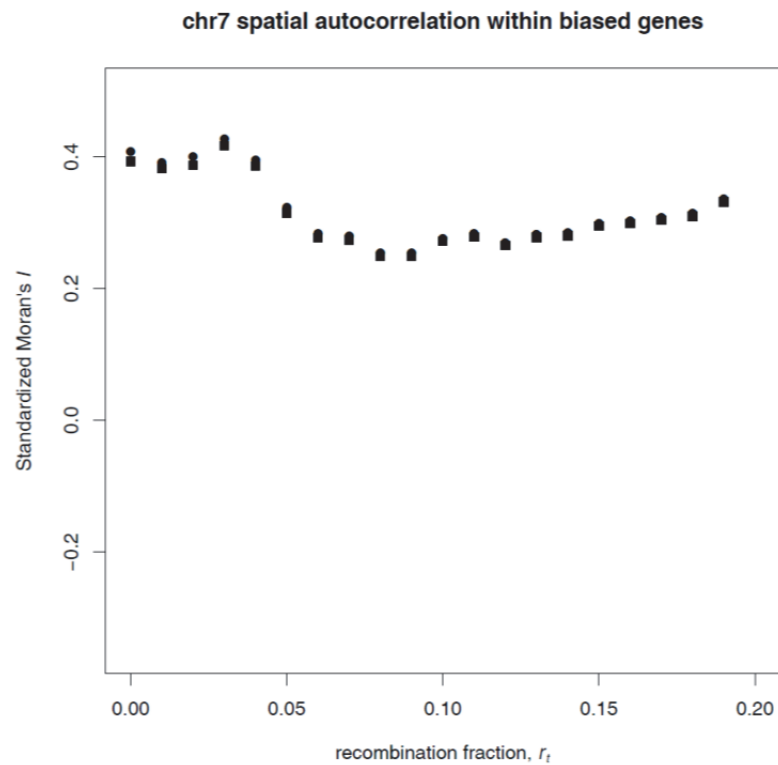


Figure 2.2 (Continued)

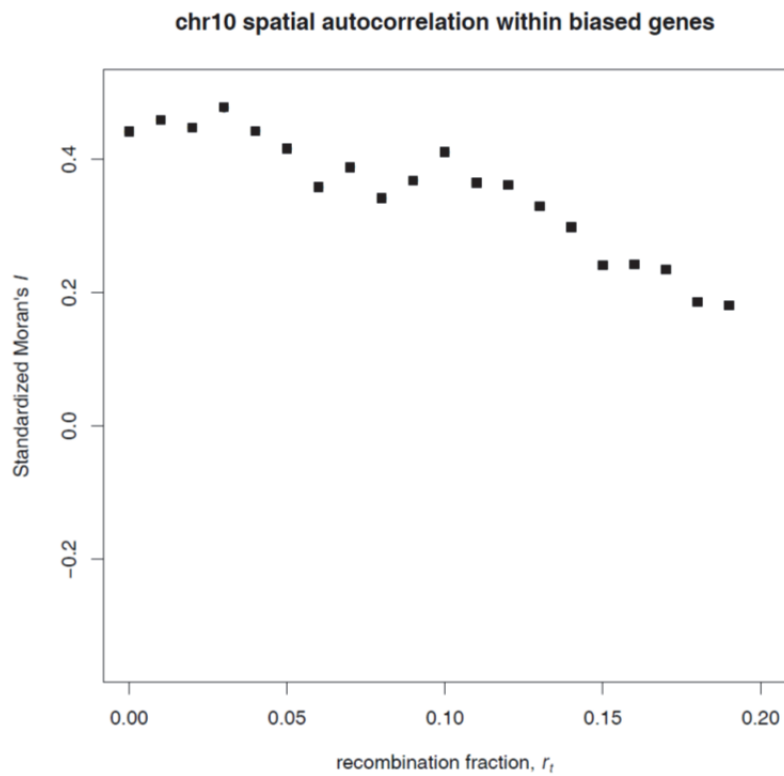
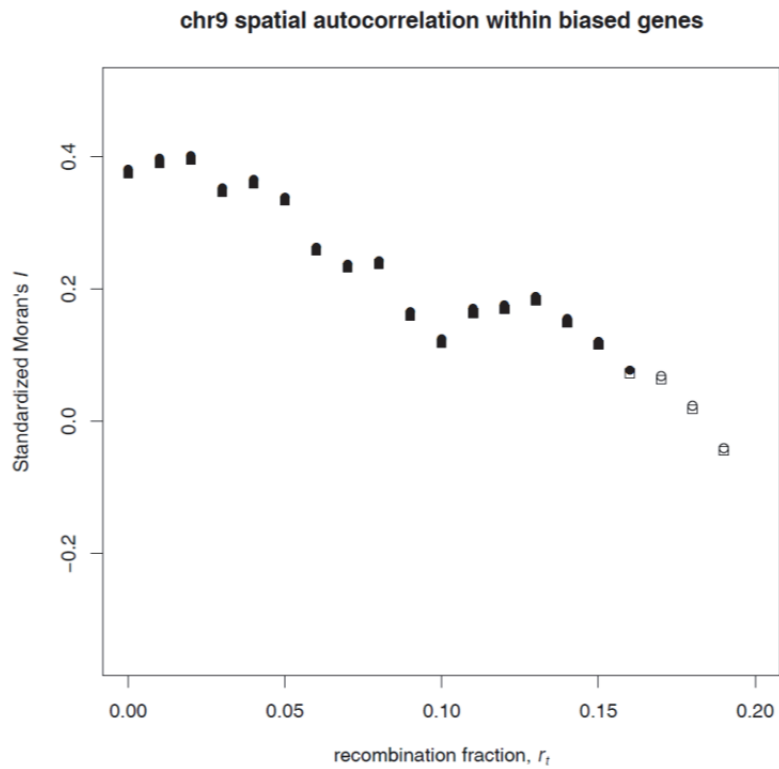


Figure 2.2 (Continued)

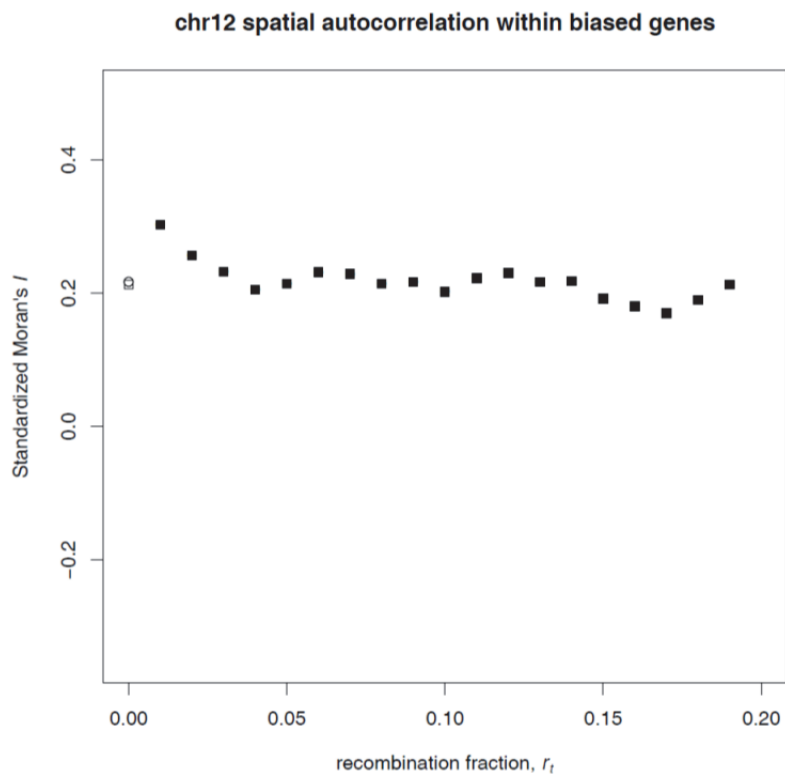
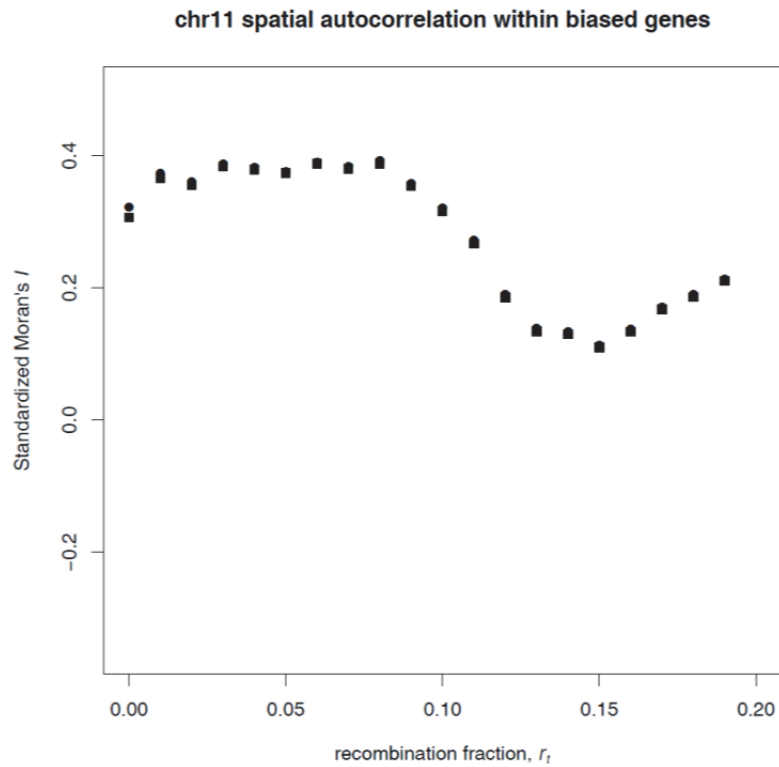


Figure 2.2 (Continued)

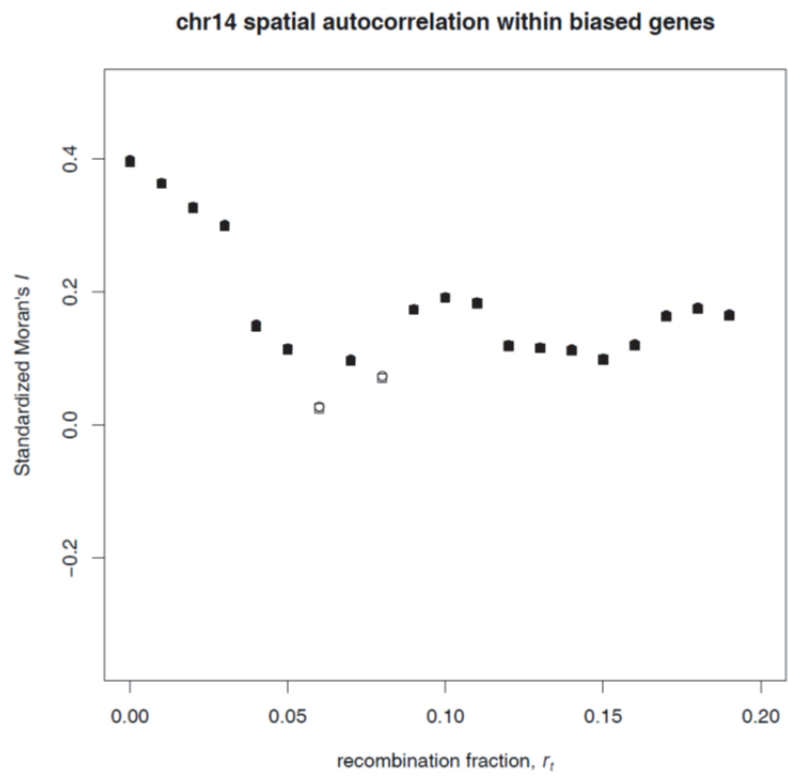
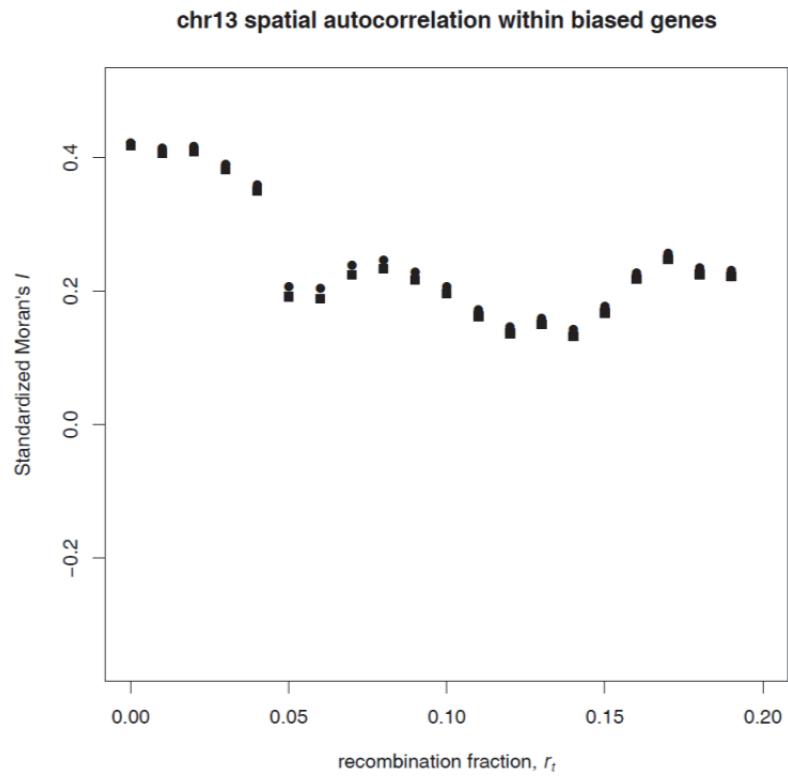


Figure 2.2 (Continued)

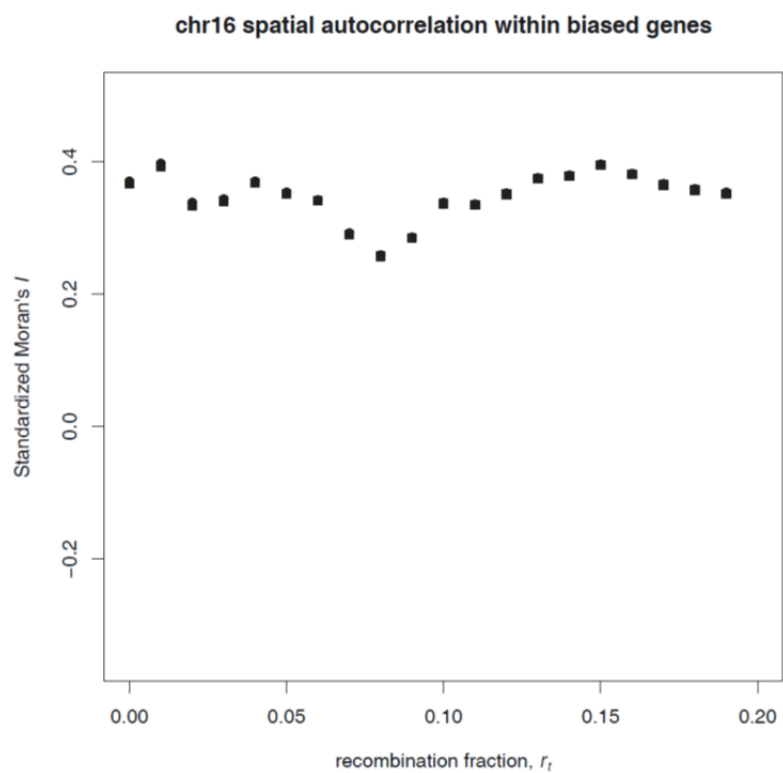
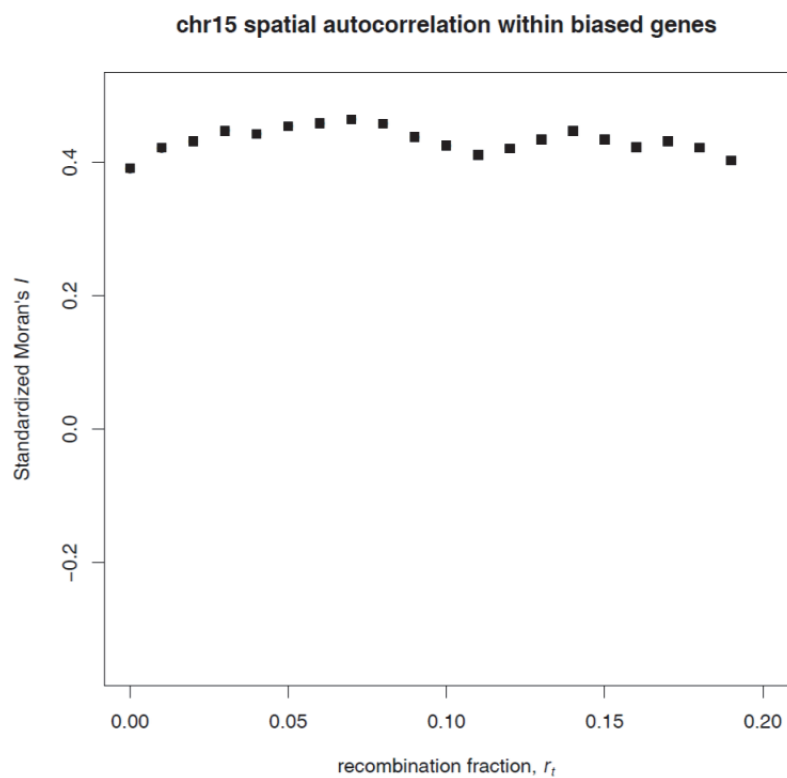


Figure 2.2 (Continued)

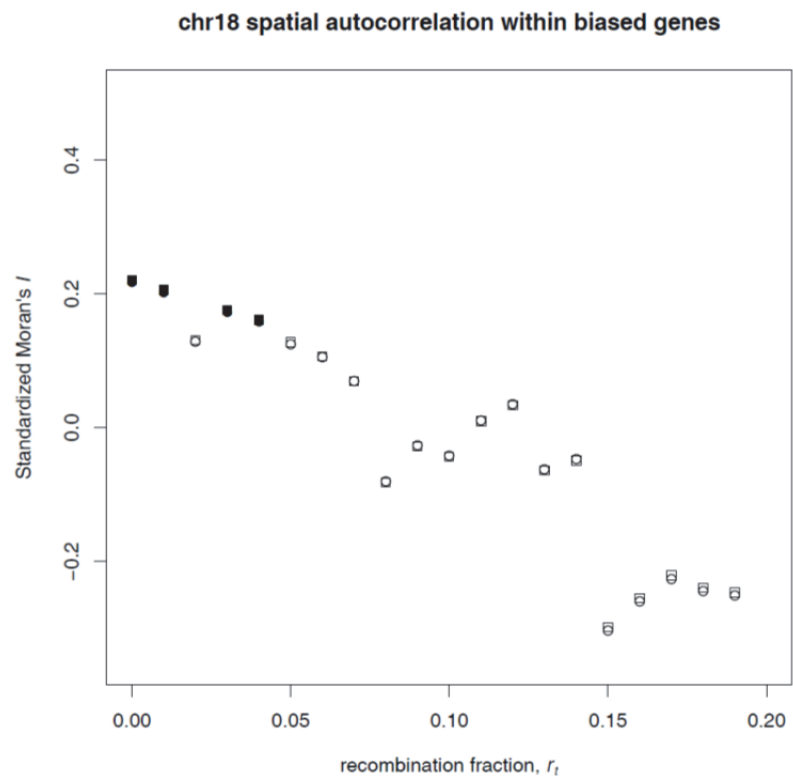
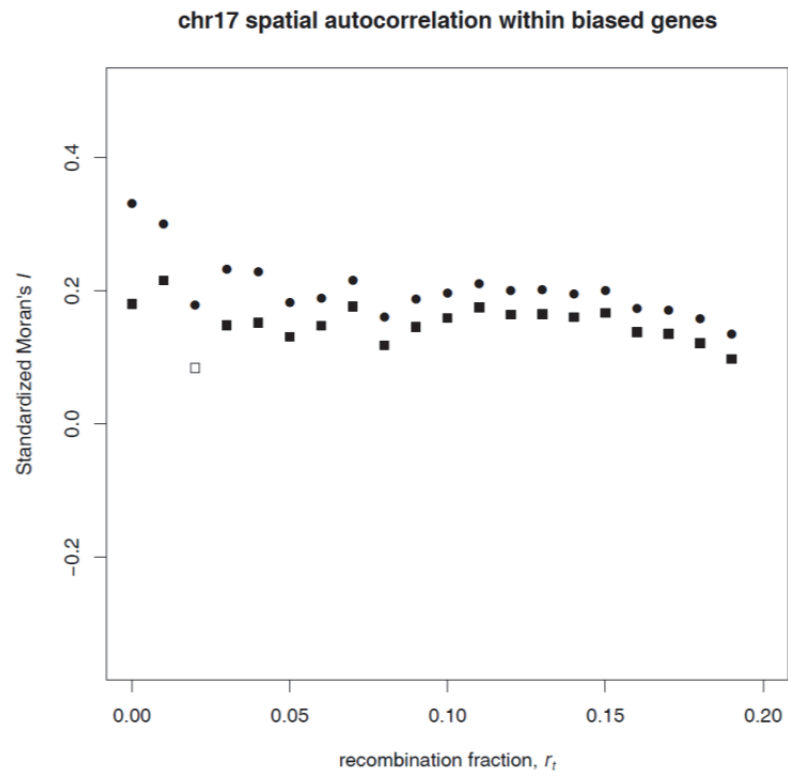


Figure 2.2 (Continued)

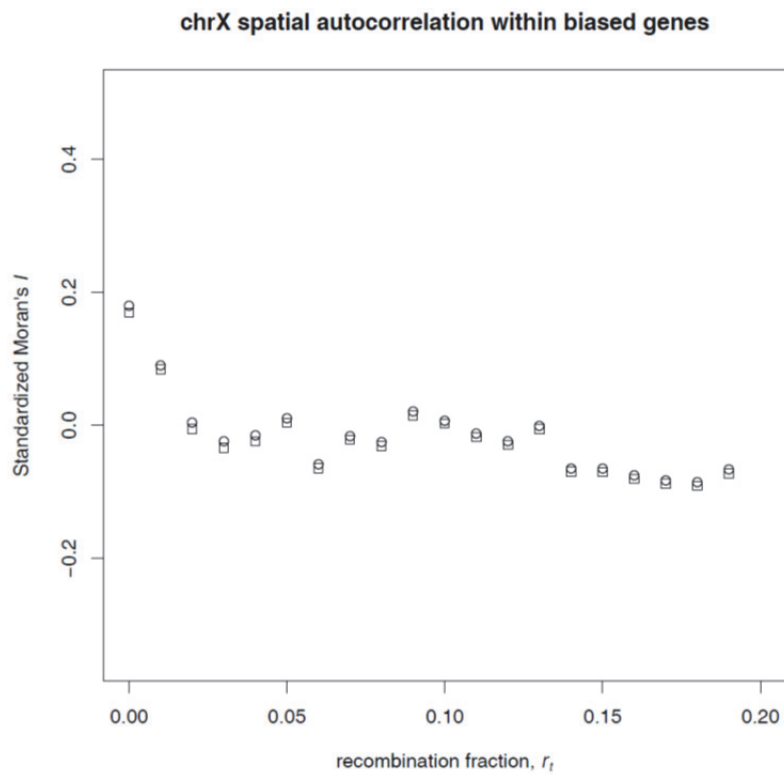
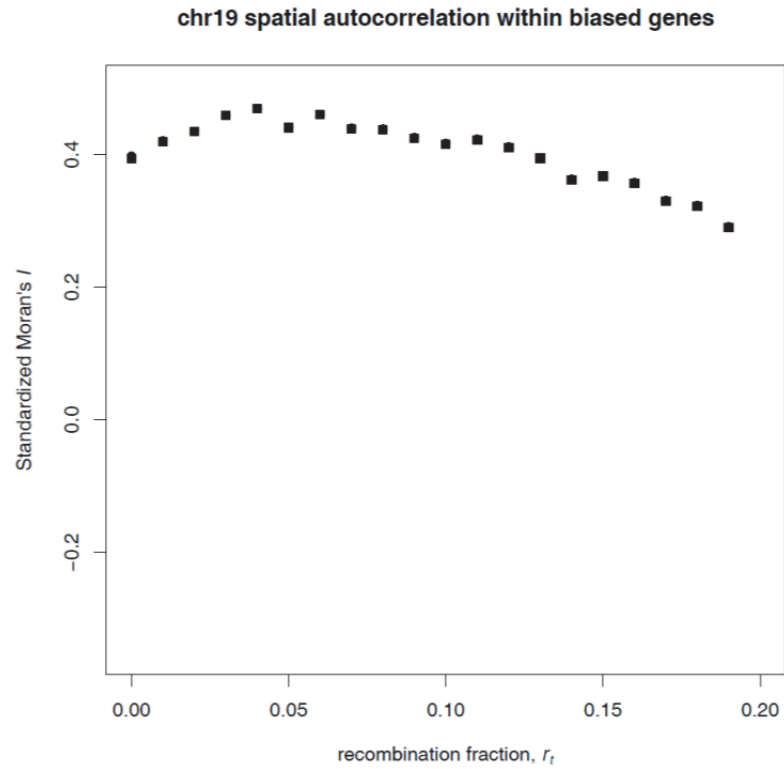


Figure 2.2 (Continued)

Maximal autocorrelation again tended to occur for somewhat low values of r_t , although this was more variable across chromosomes than in the general analysis above (Table 2.4). And again, the X chromosome was unusual in that it exhibited spatial randomness amongst its male and female-biased genes at all scales considered.

Table 2.4. Table of maximum standardized Moran's I values and their associated recombination thresholds (r_t) for the spatial autocorrelation analysis performed within the category of sex-biased genes. As described in the text, $I \in [-1, +1]$ and $r_t \in [0, 0.19]$ and the weight matrix used is given in parentheses. Positive values of I indicate clustering of like genes (e.g. male-biased genes with male-biased genes). * not significantly different from zero.

| Chromosome | I (distance) | r_t (distance) | I (distance + paralogy) | r_t (distance + paralogy) |
|--------------------------|----------------|------------------|---------------------------|-----------------------------|
| 1 | 0.425 | 0.01 | 0.422 | 0.01 |
| 2 | 0.429 | 0.02 | 0.423 | 0.02 |
| 3 | 0.445 | 0 | 0.438 | 0 |
| 4 | 0.466 | 0.04 | 0.461 | 0.04 |
| 5 | 0.423 | 0.08 | 0.421 | 0.08 |
| 6 | 0.363 | 0 | 0.358 | 0 |
| 7 | 0.428 | 0.03 | 0.417 | 0.03 |
| 8 | 0.366 | 0.07 | 0.355 | 0.07 |
| 9 | 0.402 | 0.02 | 0.395 | 0.02 |
| 10 | 0.478 | 0.03 | 0.478 | 0.03 |
| 11 | 0.392 | 0.08 | 0.388 | 0.08 |
| 12 | 0.304 | 0.01 | 0.302 | 0.01 |
| 13 | 0.422 | 0 | 0.418 | 0 |
| 14 | 0.399 | 0 | 0.395 | 0 |
| 15 | 0.465 | 0.07 | 0.464 | 0.07 |
| 16 | 0.398 | 0.01 | 0.395 | 0.15 |
| 17 | 0.332 | 0 | 0.215 | 0.01 |
| 18 | 0.217 | 0 | 0.221 | 0 |
| 19 | 0.470 | 0.04 | 0.469 | 0.04 |
| X | 0.181* | 0 | 0.169* | 0 |
| Autosomal average | 0.401 | 0.027 | 0.391 | 0.035 |

Whole chromosome organization

Finally, it is worth noting that modest autocorrelation at short recombinational scales can superficially appear as random placement along chromosomes from a much wider perspective (Figure 2.3). However, this spatially restricted yet frequent clustering along chromosomes would suggest that other genes' distances to their nearest sex-biased gene are likely small. Indeed, based on the data here, over 99% percent of these genes are within 1 cM of a sex-biased gene.

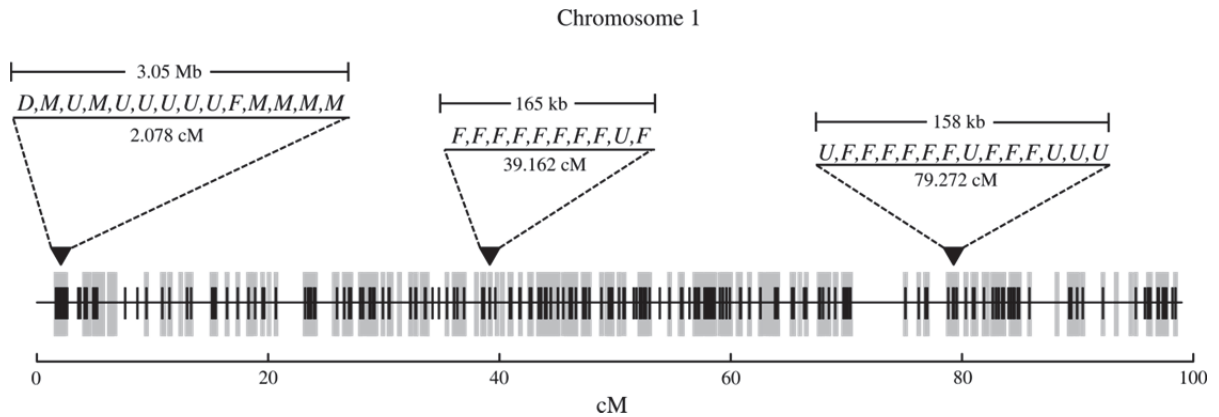


Figure 2.3. A representative depiction of the location of sex-biased genes for the entirety of chromosome 1. Sex-biased genes are shown in black and all other genes are shown in gray. As many genes can exist at a single genetic map location, three different spatially expanded regions are shown where the local positive spatial autocorrelation is particularly high. Single-letter codes represent the kind of gene located at the map position given underneath, with the physical span (i.e. in base pairs) of each region shown above. D, doubly-biased; F, female-biased; M, male-biased; U, unbiased.

Discussion

Evolutionary forces contributing to clustering

Sexually antagonistic selection is an important evolutionary process expected to influence the identity, number, chromosomal location, and distribution of many genes in the genome (Rice 1984; Patten and Haig 2009; Connallon and Clark 2010; Fry 2010; Patten et al. 2010; Connallon and Clark 2011; Frank and Crespi 2011; Úbeda et al. 2011). Recent theoretical work specifically has shown that independent but closely linked antagonistic loci maintain linkage disequilibrium

and fitness variation in excess of single locus predictions (Patten et al. 2010; Úbeda et al. 2011). These conditions in turn provide the opportunity for natural selection to resolve this conflict, most likely through the evolution of sex-biased gene expression directly or as a final step following the duplication and divergence of antagonistic loci (Bonduriansky and Chenoweth 2009). Ultimately, we may then expect to find highly nonrandom spatial relationships between sex-biased genes in real genomes. Consistent with these predictions, we find significant spatial clustering of sex-biased genes within much of the mouse genome. Notably, these patterns are strongest at the shortest spatial scales and often decline precipitously with increasing recombinational distance. Hence, the clustering of sex-biased genes is a typically localized, but frequently observed phenomenon across the majority of chromosomes, highlighting an important role for sexual antagonism in overall genome organization.

The notable exception to this genome-wide pattern of clustered sex-biased genes is observed on the X chromosome. Two factors in particular may influence this pattern. First, given male hemizyosity and female-limited recombination across most of the mammalian X, antagonistic loci any distance apart in females would be fully linked in males. Intuitively, this might reduce or eliminate the possibility of observing the sort of spatial structuring predicted among the autosomes. Second, depending on the nature and extent of epistasis between X-linked antagonistic loci and modifiers of their expression, conflict-resolving sex-biased expression may evolve fast enough to prevent spatial clustering (Connallon and Clark 2010). In line with this intuition and previous theoretical results, we observe mostly absent to weak clustering on the X at spatial scales that push the boundary for what could credibly be considered ‘close.’ Additionally, this result should be largely independent of other phenomena like X-inactivation. Estimates of genes escaping X-inactivation in *Mus* suggest they are very few in number and not

spatially clustered, making them unlikely to represent a major source of bias or confounding in this study (Yang et al. 2010). Together, the X chromosome therefore appears to be relatively unexceptional, at least along these particular dimensions of genome organization.

However, sexual antagonism can affect the structure of the X chromosome in ways other than spatial clustering. Specifically, recent theoretical work has shown that the X should be enriched for female-beneficial gene duplicates (Connallon and Clark 2011). Although analyses here do not directly speak to this prediction, they do address potential interactions between the history of gene duplications experienced by paralogous genes and the overall appearance of spatial clustering along chromosomes.

Paralogy appears to have contributed little to the clustering patterns observed here and possibly equally little to the overall history of sexual antagonism. Tandem duplication of sex-biased genes could conceivably give rise to positive spatial autocorrelation within chromosomes. However, this process had only marginal effects on our overall results. A comparison of analyses that did and did not take pair-wise paralogous relationships into account (Tables 2.3, 2.4) reveals only subtle quantitative differences in both the magnitude and spatial extent of autocorrelation across the genome. These results may downplay the role of gene duplication in sexual antagonism more generally, although we must first consider an important caveat. In our analyses, we do not actually know the relative history of sexual antagonism and gene duplication. Lacking such a precise history, it is difficult to say whether an observation of sex-biased paralogs resulted from a duplication event either prior to or in response to an invasion of sexually antagonistic alleles. However, to the extent that gene duplication actually is an important aspect of sexual antagonism, our data may instead suggest more of a role for retrotransposition over tandem duplication. Retrotransposition inserts genes into effectively

random genomic locations, and therefore would likely detract from a signal of positive spatial autocorrelation. In this way, retrotransposition could be one contributing factor towards our observed signals' relatively modest values. However, the relative rates of other processes besides gene duplication can also influence the magnitude of spatial autocorrelations observed here.

Considering our original two-locus scenario, clustering emerges only among genes presently experiencing sexual antagonism. Should this conflict be resolved for an isolated antagonistic locus, the benefits of close linkage would disappear for other antagonistic alleles invading at nearby loci. Therefore, in order to observe some degree of clustering, the waiting times for new sexually antagonistic mutations must generally be less than the rate at which previous antagonism is resolved. However, as several processes can independently influence such resolutions, this rate likely varies considerably both across the genome and across lineages. Fixation of an antagonistic allele is by far the simplest route for conflict resolution, and this can take place by a number of means (Albert and Otto 2005; Stewart et al. 2010; Connallon and Clark 2011; Connallon and Clark 2012). While such alleles remain at intermediate frequencies, however, conflict resolution can occur via a typically slower process of gene duplication and/or by the potentially rapid evolution of sex-biased expression at a locus (Stewart et al. 2010; Connallon and Clark 2011). Therefore, depending on the collective effect of these processes, portions of different chromosomes will be more or less amenable for the emergence of clustering, providing some explanation for the modest degree of spatial autocorrelation we observe.

One aspect of this spatial autocorrelation that remains unexplained, however, is the pattern observed among male and female-biased genes. Although the general pattern among all

genes is consistent with theoretical predictions, such predictions are agnostic towards potential spatial relationships *within* the subset of sex-biased genes. The model of two, independent sexually antagonistic loci (Patten et al. 2010; Úbeda et al. 2011) predicts an excess coupling of alleles that benefit the same sex, but it does not immediately follow that such linked loci would evolve the same bias in *expression* as well. For example, higher expression may benefit males at one locus, whereas lower expression may benefit males at a second locus. Over time, selection would then respectively favor male and female-biased expression independently at these two loci. Given our general ignorance on the actual mapping of gene expression to fitness, we may reasonably only expect spatial randomness *within* sex-biased genes at a genome-wide scale. However, that we can reject such randomness across all autosomes with the data here (Table 2.4) suggests that other factors must be at play. Contrary to the assumption of the two-locus model, not all antagonistic genes are likely to influence fitness independently. Specifically, there may actually be considerable epistasis between linked antagonistic genes and this in turn might favor a coordination of their respective sex-biases, giving rise to the pattern observed here. Although this possibility deserves proper theoretical treatment to test its validity in the context of sexual antagonism, similar logic has recently proven true for epistatically interacting clusters of imprinted genes (Wolf 2013).

Future implications of clustering

A variety of processes have contributed to the observed organization of the *Mus* genome. However, given this degree of clustering, what are its potential consequences for future sexually antagonistic selection? We argue that this genome should represent a fairly permissive

environment for both current and future segregating sexually antagonistic alleles. Although the generally restricted spatial nature of the clustering we have described might suggest otherwise, the superficially even looking distribution of sex-biased genes along chromosomes (Figure 2.3) contributes favorably towards this environment. Specifically, no gene is all that far from its nearest sex-biased gene in terms of recombinational distances. Provided these sex-biased genes represent incompletely resolved conflict, sufficient fitness variation may remain to facilitate the invasion and maintenance of new antagonistic alleles at other closely linked loci. In this way, the spatial arrangement of genes in conflict can make the entire genome more amenable to the evolution of future conflict.

As stable conflict leaves both sexes in a maladapted state, it may seem that having a genome prone to maintaining further conflict would only exacerbate this issue. However, depending on the mode and rate of conflict resolution, the exact opposite result may counterintuitively be true. Stably maintained antagonism at a single locus or set of linked loci maintains fitness variation over time. Stable fitness variation thus translates into constant selective pressure acting on other mutations that serve to resolve this conflict. Whether through gene duplication or sex-biased expression, these conflict-resolving mutations effectively assist in genetically decoupling the sexes' shared phenotypes from one another. This decoupling would thus reduce the cross-sex genetic correlation (Lande 1980; Bonduriansky and Chenoweth 2009; van Doorn 2009), facilitating evolution towards sex-specific optima at an increased rate. In this way, clustered genomes effective at maintaining conflict in the short-term may be similarly effective at ultimately overcoming it in the long-term to achieve sex-specific adaptation. From this perspective, genome organization is relevant not only to our proximate understanding of

genotype-phenotype maps, but also to our understanding of large-scale patterns and the dynamics of phenotypic evolution among sexual taxa.

Acknowledgments

This chapter benefitted greatly from discussions with Manus Patten, Edgar Duéñez-Guzmán, and Mark Cornwall on the underlying theory and computational approaches for addressing the resulting predictions. I am also grateful to Sarah Kocher for commentary and suggestions on previous versions of this chapter and to the Department of Organismic and Evolutionary Biology for financial support over the past several years.

References

- Albert, A. Y. K. and S. P. Otto. 2005. Sexual selection can resolve sex-linked sexual antagonism. *Science* 310: 119-121.
- Auer, P. L. and R. W. Doerge. 2010. Statistical design and analysis of RNA sequencing data. *Genetics* 185: 405-416.
- Bonduriansky, R. and S. F. Chenoweth. 2009. Intralocus sexual conflict. *Trends in Ecology & Evolution* 24: 280-288.
- Cliff, A. D., P. Haggett, J. K. Ord, K. A. Bassett and R. B. Davies. 1975. *Elements of spatial structure: a quantitative approach*. Cambridge University Press, Cambridge, U.K.
- Cliff, A. D. and J. K. Ord. 1973. *Spatial autocorrelation*. Pion Limited, London, U.K.
- Connallon, T. and A. G. Clark. 2010. Sex linkage, sex-specific selection, and the role of recombination in the evolution of sexually dimorphic gene expression. *Evolution* 64: 3417-3442.
- Connallon, T. and A. G. Clark. 2011. The resolution of sexual antagonism by gene duplication. *Genetics* 187: 919-937.
- Connallon, T. and A. G. Clark. 2012. A general population genetic framework for antagonistic selection that accounts for demography and recurrent mutation. *Genetics* 190: 1477-1489.
- Cox, A., C. L. Ackert-Bicknell, B. L. Dumont, Y. Ding, J. T. Bell, G. A. Brockmann, J. E. Wergedal, C. Bult, B. Paigen, J. Flint, S. W. Tsaih, G. A. Churchill and K. W. Broman. 2009. A new standard genetic map for the laboratory mouse. *Genetics* 182: 1335-1344.
- Dawkins, R. 1976. *The selfish gene*. Oxford University Press, Oxford, U.K.
- Ellegren, H. and J. Parsch. 2007. The evolution of sex-biased genes and sex-biased gene expression. *Nature Reviews Genetics* 8: 689-698.
- Frank, S. A. and B. J. Crespi. 2011. Pathology from evolutionary conflict, with a theory of X chromosome versus autosome conflict over sexually antagonistic traits. *Proceedings of the National Academy of Sciences* 108: 10886-10893.
- Fry, J. D. 2010. The genomic location of sexually antagonistic variation: some cautionary comments. *Evolution* 64: 1510-1516.
- Fujita, P. A., B. Rhead, A. S. Zweig, A. S. Hinrichs, D. Karolchik, M. S. Cline, M. Goldman, G. P. Barber, H. Clawson, A. Coelho, M. Diekhans, T. R. Dreszer, B. M. Giardine, R. A. Harte, J. Hillman-Jackson, F. Hsu, V. Kirkup, R. M. Kuhn, K. Learned, C. H. Li, L. R. Meyer, A. Pohl, B. J. Raney, K. R. Rosenbloom, K. E. Smith, D. Haussler and W. J. Kent. 2011. The UCSC Genome Browser database: update 2011. *Nucleic Acids Research* 39: D876-D882.

- Gallach, M., C. Chandrasekaran and E. Betran. 2010. Analyses of nuclearly encoded mitochondrial genes suggest gene duplication as a mechanism for resolving intralocus sexually antagonistic conflict in *Drosophila*. *Genome Biology and Evolution* 2: 835-850.
- Gibson, J. R., A. K. Chippindale and W. R. Rice. 2002. The X chromosome is a hot spot for sexually antagonistic fitness variation. *Proceedings of the Royal Society B: Biological Sciences* 269: 499-505.
- Gregg, C., J. W. Zhang, B. Weissbourd, S. J. Luo, G. P. Schroth, D. Haig and C. Dulac. 2010. High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science* 329: 643-648.
- Haig, D. 2012. The strategic gene. *Biology and Philosophy* 27: 461-479.
- Hurst, L. D., C. Pal and M. J. Lercher. 2004. The evolutionary dynamics of eukaryotic gene order. *Nature Reviews Genetics* 5: 299-310.
- Kodama, Y., M. Shumway and R. Leinonen. 2012. The sequence read archive: explosive growth of sequencing data. *Nucleic Acids Research* 40: D54-D56.
- Kosambi, D. 1943. The estimation of map distances from recombination values. *Annals of Eugenics* 12: 172-175.
- Lande, R. 1980. Sexual dimorphism, sexual selection, and adaptation in polygenic characters. *Evolution* 34: 292-305.
- Lercher, M. J. and L. D. Hurst. 2006. Co-expressed yeast genes cluster over a long range but are not regularly spaced. *Journal of Molecular Biology* 359: 825-831.
- Mank, J. E., L. Hultin-Rosenberg, M. T. Webster and H. Ellegren. 2008. The unique genomic properties of sex-biased genes: Insights from avian microarray data. *BMC Genomics* 9: 148.
- Moran, P. A. 1950. Notes on continuous stochastic phenomena. *Biometrika* 37: 17-23.
- Moran, P. A. P. 1948. Some theorems on time series: II the significance of the serial correlation coefficient. *Biometrika* 35: 255-260.
- Patten, M. M. and D. Haig. 2009. Maintenance or loss of genetic variation under sexual and parental antagonism at a sex-linked locus. *Evolution* 63: 2888-2895.
- Patten, M. M., D. Haig and F. Úbeda. 2010. Fitness variation due to sexual antagonism and linkage disequilibrium. *Evolution* 64: 3638-3642.
- Rice, W. R. 1984. Sex chromosomes and the evolution of sexual dimorphism. *Evolution* 38: 735-742.

- Semon, M. and L. Duret. 2006. Evolutionary origin and maintenance of coexpressed gene clusters in mammals. *Molecular Biology and Evolution* 23: 1715-1723.
- Stewart, A. D., A. Pischedda and W. R. Rice. 2010. Resolving intralocus sexual conflict: genetic mechanisms and time frame. *Journal of Heredity*. 101: S94-S99.
- Storey, J. D. and R. Tibshirani. 2003. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences* 100: 9440-9445.
- Trapnell, C., L. Pachter and S. L. Salzberg. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25: 1105-1111.
- Úbeda, F., D. Haig and M. M. Patten. 2011. Stable linkage disequilibrium owing to sexual antagonism. *Proceedings of the Royal Society B: Biological Sciences* 278: 855-862.
- Upton, G. J. G. and B. Fingleton. 1985. *Spatial data analysis by example*. John Wiley & Sons, New York.
- van Doorn, G. S. 2009. Intralocus sexual conflict. *Annals of the New York Academy of Sciences* 1168: 52-71.
- Vilella, A. J., J. Severin, A. Ureta-Vidal, L. Heng, R. Durbin and E. Birney. 2009. EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Research* 19: 327-335.
- Weber, C. C. and L. D. Hurst. 2011. Support for multiple classes of local expression clusters in *Drosophila melanogaster*, but no evidence for gene order conservation. *Genome Biology* 12: R23.
- Williams, G. C. 1966. *Adaptation and natural selection: a critique of some current evolutionary thought*. Princeton University Press, Princeton, NJ.
- Wolf, J. B. 2013. Evolution of genomic imprinting as a coordinator of coadapted gene expression. *Proceedings of the National Academy of Sciences* 110: 5085-5090.
- Yang, F., T. Babak, J. Shendure and C. M. Disteche. 2010. Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Research* 20: 614-622.
- Yang, X., E. E. Schadt, S. Wang, H. Wang, A. P. Arnold, L. Ingram-Drake, T. A. Drake and A. J. Lusis. 2006. Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Research* 16: 995-1004.

CHAPTER 3

Hormones, sexual dimorphism, and the resolution of sexual antagonism

Abstract

Sexual antagonism over the value of a shared trait favors the evolution of sexual dimorphism. Many dimorphic whole-organism phenotypes are determined in part by the actions of steroid sex hormones. Intuition would suggest and experimental evidence confirms that sex hormones strongly influence dimorphism at the level of individual gene expression as well. However, many discussions of sex-biased gene expression make implicit assumptions on the relationship between hormones and dimorphism that remain largely unexamined. Using genome-wide expression data from *Mus*, we examined the association between sex-biased gene expression and presence/identity of *cis*-acting promoter hormone response elements. This association was surprisingly weak, with the majority of sex-biased genes lacking any hormone response elements. This result implies a greater role for *trans*-acting factors on sex-biased expression than was previously appreciated. Additionally, we find the potential for new conflict arising from within this *trans*-regulatory environment, due in part to divergent interests between X-linked and autosomal genes. Together, these results suggest that although sexual dimorphism at a locus may evolve quickly, the actual resolution of sexual antagonism may proceed more slowly.

Introduction

Ecological differences between the sexes can lead to sexually antagonistic selection, whereby phenotypes shared by males and females are selected in opposing directions. However, most genes spend equal time in the two sexes and this should attenuate divergence in phenotype (Lande 1980; Poissant et al. 2010). Loci responsible for such phenotypes may thus harbor alleles that benefit one sex while reducing the fitness of the other, a situation commonly known as intralocus sexual conflict (hereafter also referred to as sexual antagonism) (Rice 1984; Rice and Chippindale 2001; Bonduriansky and Chenoweth 2009; van Doorn 2009). While these alleles remain at intermediate frequencies, antagonism persists and neither sex may reach their respective fitness optimum (Rice 1984; Patten and Haig 2009; Fry 2010; Patten et al. 2010; Frank and Crespi 2011). Antagonism thus leaves a population in a semi-permanently maladapted state. In such circumstances, natural selection favors mutations that can genetically decouple male and female interests. Sex-specific adaptation could then proceed unhindered and antagonism would eventually be resolved. For genes in conflict over expression levels, mutations causing sex-biased expression of these genes represent one such target of natural selection. A focus on sex-biased genes raises a new question, however. How exactly does a gene know what sex it is in?

In order to exhibit sex-biased expression, a gene must recognize some external signal of its sexual environment. Two possible signals commonly considered are steroid sex hormones and the sex chromosomes. Sex chromosomes can act as signals of sexual identity by virtue of their inheritance patterns. Genes located on or directly regulated by the Y-chromosome (or W) have immediate information on their location within the heterogametic sex. Similarly, the haplodiploid inheritance of the X-chromosome (or Z) might contribute information to genes on

their sexual environment. But upon closer examination, the sex chromosomes are either less credible or less useful sources of information than this description implies. The Y-chromosome is both genetically depauperate and is not thought to have extensive transcriptional control over the rest of the genome (but see Lemos et al. 2008). And given the vagaries of dosage compensation, random inactivation, and its presence in both sexes, the X-chromosome may represent a more dubious source of information on a gene's sexual environment than is typically appreciated. Finally, some gonochoristic taxa lack sex chromosomes and therefore cannot rely on their presence to differentiate between the sexes. Collectively, these observations suggest that our other possibility—steroid sex hormones—may be a more informative and commonly used signal of a gene's sexual environment for the production of sex-biased expression.

The basic biology of the steroid sex hormones paints a clear and intuitively satisfying picture of their role as ubiquitous signals of sexual identity. These hormones are divided into two families, androgens and estrogens, which are respectively produced primarily from male and female gonads (Arnold 2009). In addition to their clear sex-specific origins, these hormones' actions also influence prominent phenotypic sex differences. These influences are most easily seen through the effects of mutations in sex hormone pathways. For example, XY individuals with inactivation mutations of the androgen receptor cannot respond to their own circulating androgens. This insensitivity causes the development of female external genitalia and secondary sexual characteristics (Notini et al. 2005; Galani et al. 2008). Similarly, females lacking functional estrogen receptors exhibit underdeveloped mammary glands and develop testis-specific structures within ovaries (Korach 1994; Couse et al. 1999). Emerging from these basic observations is the notion that androgens thus act as molecular signals of maleness and estrogens as signals of femaleness. How this notion relates to sexual dimorphism at the level of individual

gene expression is all the more clear when we consider the finer scale mechanisms of hormone actions.

Although effects of sex hormones are most visible at the level of whole-body phenotypes, these effects ultimately result from hormones' canonical intracellular actions. Sex hormones circulating in the bloodstream are capable of diffusing directly into cells where they bind to separate androgen and estrogen-specific receptors. These receptors also function as transcription factors. Specifically, these receptors bind to roughly 15 base pair motifs, known as hormone response elements (HREs), located near transcription start sites (Verrijdt et al. 2003; Gruber et al. 2004). While bound to these *cis*-acting elements, steroid receptors, in their role as transcription factors, promote the expression of adjacent genes. Combining this genetic mechanism with the notion of sex-specific hormonal environments, we can now arrive at a simple and intuitive framework implicit to many discussions on the evolution of sex-biased gene expression (Rhen 2000; Coyne et al. 2008; Mank 2009; Stewart et al. 2010).

The information presented above suggests a simple means by which genes can both recognize and react to their sexual environment, thus laying out a parsimonious scenario for how sex-biased expression could evolve. Circulating androgens and estrogens act as reliable signals to genes that they are located in male and female bodies, respectively. Genes can then respond to these signals by mutationally acquiring an appropriate HRE somewhere in their promoters. This scenario predicts that male-biased genes should contain an androgen response element, female-biased genes should contain an estrogen response element, and unbiased genes should lack both (Figure 3.1). Hence, sex-biased expression could evolve most simply and quickly by the independent, locus-by-locus acquisition of hormone response elements acting in *cis*.

| | Androgen sensitive | Estrogen sensitive | Hormone insensitive |
|---------------|--------------------|--------------------|---------------------|
| Male-biased | ++++ | - | - |
| Female-biased | - | ++++ | - |
| Unbiased | - | - | ++++ |

Figure 3.1. A conceptual model representing basic intuitive predictions on the connection between hormones and sex-biased expression patterns. Within this framework, we might expect genes to be enriched along the diagonal (represented by plus signs) and under-represented or entirely absent off the diagonal (represented by minus signs). Cells of this table enclosed by the dotted line constitute the model's core, with the remainder of the table representing a corollary of this logic.

To our knowledge, this intuitive scenario and its specific predictions have never been formally tested. This is unfortunate as evidence for or against this scenario could clarify some standing controversies. Specifically, arguments over the rate at which sex-biased expression evolves and the degree to which it represents a resolution of sexual antagonism are highly dependent upon the particular genetic mechanisms invoked (Stewart et al. 2010).

Therefore, we conduct a simple test of the relation between sex-biased expression and possession of HREs while also considering other factors relevant to conflict resolution. Specifically, we estimate the degree of influence of *trans*-acting genes on patterns of sex-biased expression. The framework described above assumes that all genetic regulation acts in *cis*, yet *trans*-acting genes can also serve as possibly independent signals of their target genes' sexual

environment. Additionally, assessing the relationships between these factors and their chromosomal context is crucial for understanding conflict resolution. Non-overlapping evolutionary interests between X-linked and autosomal genes could considerably slow conflict resolution, yet this phenomenon has received relatively little attention (Frank and Crespi 2011). By considering this interplay between genes' hormonal, *trans*-regulatory, and chromosomal environments, we therefore attempt to gauge the explanatory power of our current genomic intuition on sex-biased expression and the resolution of sexual antagonism.

Methods

Gene expression data

To address the predictions of our model of hormone-based sex-biased expression (Figure 3.1), we first used the same data underlying analyses in Chapter 2 to produce our estimates of sex-biased expression here. However, unlike in Chapter 2, this sample of genes was not constrained by the limited extent of linkage map coverage. Therefore, we were able to estimate sex-biased expression in a slightly larger set of genes in the genome. We again used q -value and expression bias cutoffs to control the false discovery rate and diagnose significant sex-biased expression. Within this sample, a gene was considered unbiased unless $q < 0.01$ and the observed expression bias was greater than 1.2.

Hormone response elements

We searched each gene's promoter for hormone response elements (HREs) in a 15 kilobase window (−10 to +5 kb surrounding the transcription start site) for the canonical

transcript of each gene in the mm9 genome. We chose this particular window as it has been the focus of previous studies and because it represents a functionally relevant position for known and active HREs measured experimentally (Bourdeau et al. 2004). Within this window, we searched for canonical and non-canonical androgen and estrogen response elements (AREs, EREs respectively) using a custom Python script. Presence or absence of at least one ARE or ERE was scored for each gene (Table 3.1; Claessens et al. 2001; Verrijdt et al. 2003; Bourdeau et al. 2004; Gruber et al. 2004; Moehren et al. 2008).

Table 3.1. Hormone response element regular expressions. The following are the somewhat inelegant regular expression strings used to represent a variety of canonical and non-canonical androgen and estrogen response elements (ARE, ERE). Regular expressions contain both entire motifs as well as individually variable positions within those motifs.

| Hormone response element | Regular expression string | References |
|--------------------------|--|---|
| ARE | 'GGTACA...TGTTCT GTTACA...TGTTCT AGTACG...TGTTCT AGCACT...TGTTCT GAAACA...TGTTCT AGTACT...TGTTCT AGCACG...GGTTCC TGAA GT...TGTTCT AGCACT...TGTTCC AATACA...TGTTCC GGTTCT...AGTACT GGCTCT...AGTTCT TGGTCA...AGTTCT' | Claessens et al. 2001; Moehren et al 2008; Verrijdt et al. 2003 |
| ERE | '[AG]GGTC[AG]...[TC][CG]ACC[TC] GGTCA...TGACC GGTCA...TGGCC GGCA...TGACC GGCCG...TGACC GGGTA...TGACC GGTGG...TGACC AATCA...TGACT AGGTCA.{10,25}AGGTCA TGACCT.{10,25}TGACCT' | Bourdeau et al. 2004; Gruber et al. 2004 |

Trans-actions

To capture a genome-wide view of the *trans*-regulatory environment, we used the Gene Ontology-based categorization scheme of the Animal Transcription Factor Database to arrange genes into four functional categories: transcription factors (TFs), transcription cofactors, chromatin remodelers (Zhang et al. 2012), and “other.” For a subset of analyses, we collapsed the first three categories into the meta-category “expression-related genes.”

Next, we summarized information on some of the targets of expression-related genes to assess the impact of the *trans*-regulatory environment on downstream expression patterns. We gathered lists from the Animal Transcription Factor Database of known or estimated target genes regulated by some of the TFs expressed in our entire dataset. Subsets of these lists overlapped with the set of all genes expressed here, thus allowing us to categorize the expression bias of these targets (male, female, or unbiased). For each TF, we then summarized the average bias of its target genes. This would then allow us to measure the concordance between the sex-bias of TFs and their target genes. To accomplish this, we recoded the categorical measure of target genes’ sex-bias into a numerical measure. Specifically, female-biased targets were coded as -1 , unbiased targets as 0 and male-biased targets as $+1$. For each TF individually, we then calculated the arithmetic mean of the values of each of their target genes. The resulting average target bias index is thus a continuous measure that ranges from -1 to $+1$. TFs with these boundary values would thus regulate all female-biased or all male-biased target genes, respectively.

Statistical analyses

Statistical analyses were performed in R v. 2.14.1 (R Development Core Team 2011). Associations between categorical variables (e.g. expression bias, HRE, chromosome, molecular function) were assessed by Pearson's chi-square tests. Standardized residuals for the cells of these tables were also calculated following Agresti (2007). Standardized residuals with an absolute value greater than 2 were taken to indicate a cell's lack of fit to the null hypothesis of no association between variables. However, these values are only reported for tables with an overall chi-square *p*-value of less than 0.05. Finally, nonparametric Wilcoxon rank sum tests were used to analyze our sole continuous variable, the average target bias index.

Results

Biased gene expression and hormone response elements

We directly estimated the expression bias of 17,851 genes in the MPFC and POA (62.9% of all properly localized autosomal and X-linked genes). 93.4% were expressed in both brain regions and we estimated their bias separately in each. *Q*-value and threshold-corrected Fisher's exact test results detected 1,782 biased genes within the POA and 6,844 biased genes within the MPFC. Genes biased in only one or both tissues were considered biased in subsequent analyses, with the exception of the few genes that exhibited opposite biases in the two brain regions. These oppositely biased genes represented only 1.2% of the total and were removed from all subsequent analyses to avoid counting each gene twice. We were thus left with a total of 17,643 genes; 4,097 of these genes were female-biased, and 3,374 were male-biased.

A regular expression search of the 15 kilobase window surrounding the transcription start site indicated that approximately 37% of these genes contained at least one hormone response

element (HRE), including 3.6% with both androgen and estrogen response elements. However, to align our data with the structure of our model (Figure 3.1), we removed these genes from subsequent analyses involving the HRE categorical variable.

Expression bias associations

Biased gene expression was significantly associated with the presence of a putative HRE ($\chi^2 = 71.8$, $df = 4$, $P < 0.001$) but the association was weak, with more than 70% of male-biased genes and 62% of female-biased genes lacking HREs (Table 3.2).

Table 3.2. Observed counts of genes with expression bias and hormone response element categories. Standardized residuals are given in parentheses for each cell only for variables exhibiting a significant association (chi-square p -value < 0.05). Residuals with an absolute value greater than 2 are shown in bold to signify significant departures from the null hypothesis for a particular cell. The table is also arranged to match the structure of Figure 3.1 to facilitate comparison of these results with intuitive predictions.

| All genes | ARE | ERE | None |
|---------------------------------|----------------------|----------------------|-----------------------|
| Male biased | 315 (-2.05) | 642 (-6.16) | 2322 (+6.84) |
| Female biased | 395 (-1.07) | 1068 (+6.2) | 2435 (-4.87) |
| Unbiased | 1092 (2.54) | 2319 (-0.36) | 6420 (-1.33) |
| | | | |
| Expression-related genes | ARE | ERE | None |
| Male biased | 39 | 75 | 272 |
| Female biased | 23 | 77 | 214 |
| Unbiased | 87 | 173 | 499 |

Female-biased genes were significantly over-represented, and male-biased genes under-represented, among genes containing an ERE. Nevertheless, 20% of all male-biased genes possessed an ERE. Unbiased genes were significantly over-represented and male-biased genes were under-represented among genes containing an ARE. Therefore, the distribution of EREs was consistent with the model's predictions but the distribution of AREs directly contradicted the model.

We next considered whether sex-biased genes could be disproportionately associated with chromosomal and *trans*-regulatory environments. Similar to other studies and in line with theoretical predictions (Rice 1984; Wang et al. 2001; Parisi et al. 2003), we found a significant association between gene expression bias and autosomal versus X-linkage ($\chi^2 = 166.75$, $df = 2$, $P < 0.001$; Table 3.3). Specifically, the X chromosome is enriched for male-biased genes but depauperate for female-biased genes. Considering *trans*-environments, we found that over-represented male-biased genes drive an association between expression bias and expression-related molecular function ($\chi^2 = 56.19$, $df = 2$, $P < 0.001$; Table 3.3). Taking both chromosomal and *trans*-environments into account, we then investigated the association between expression bias and chromosomal location solely for expression-related genes. These variables remained significantly associated ($\chi^2 = 14.31$, $df = 2$, $P < 0.001$; Table 3.3), although female-biased genes were no longer under-represented on the X chromosome. Finally, it is worth noting that, contrary to the total gene set, these *trans*-acting genes exhibited no association between their expression bias and HRE category ($\chi^2 = 6.8$, $df = 4$, $P = 0.146$; Table 3.2).

Table 3.3. Observed counts of genes with expression bias and chromosomal or functional categories. Analyses making use of the chromosomal category are shown for both the entire set of genes and for the subset of solely expression-related genes. Standardized residuals are presented as in Table 3.2.

| All genes | Autosome | X-chromosome |
|---------------------------------|-----------------------|-----------------------|
| Male biased | 3139 (-12.7) | 235 (+12.7) |
| Female biased | 4015 (+5.64) | 82 (-5.64) |
| Unbiased | 9889 (+5.29) | 283 (-5.29) |
| | | |
| Expression-related genes | Autosome | X-chromosome |
| Male biased | 376 (-3.66) | 24 (+3.66) |
| Female biased | 323 (+0.26) | 10 (-0.26) |
| Unbiased | 772 (+3.01) | 15 (-3.01) |
| | | |
| All genes | Expression-related | Other |
| Male biased | 400 (+7.46) | 2974 (-7.46) |
| Female biased | 333 (-1.27) | 3764 (+1.27) |
| Unbiased | 787 (-4.85) | 9385 (+4.85) |

Biased transcription factors and their downstream effects

To estimate the influence of the *trans*-regulatory environment on sex-biased expression, we measured the agreement between the sex-bias of transcription factors (TFs) and that of their downstream targets. Specifically, we tested the hypothesis that a sex-biased TF should impart its bias to its downstream targets. Under this hypothesis, we thus expected that male and female-biased TFs' average target bias indices should skew towards +1 and -1, respectively. This test made use of target gene information available for 44 of the 547 sex-biased TFs measured here. These 44 TFs collectively regulate 529 target genes whose expression bias was also measured here.

Comparing the distributions of the average target bias index for male and female-biased TFs, our hypothesis was weakly supported by the visual trend of the data (Figure 3.2). However, this difference between male and female-biased TFs did not quite reach significance (Wilcoxon rank sum test, $W = 161.5$, 1-tailed $P = 0.051$), due in part to considerable variation within the two groups. One contributing factor to this variation is the multifaceted regulation of target genes. Specifically, 188 of the 529 target genes are regulated by more than one biased TF (range: 2-9), with 114 of these targets being regulated by at least one male *and* one female-biased TF. These counts are likely underestimates however, as the data here represent a small fraction of not only the total pool of TFs, but also the subset of sex-biased TFs.

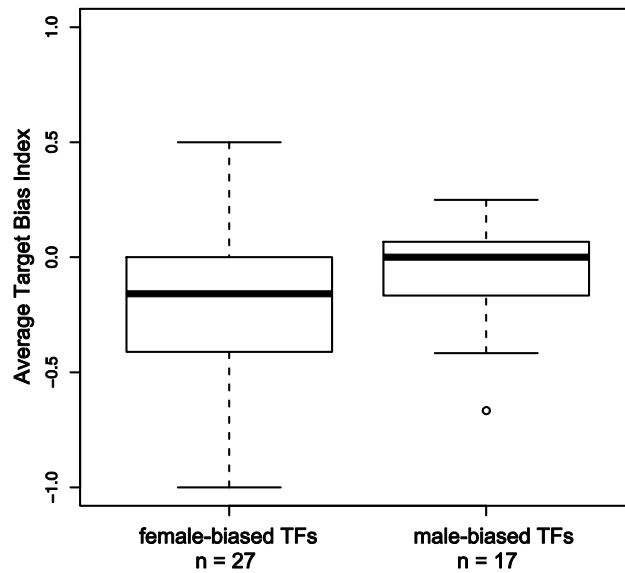


Figure 3.2. Box plots representing the distributions of average target bias index values for female and male-biased transcription factors (TFs) separately. A transcription factor with all female or all male-biased target genes would have a value of this index of -1 or $+1$, respectively.

Discussion

Steroid hormones play a well-established role in the development and physiology of major phenotypic differences between the sexes. The logic of hormonal influences on whole-organism phenotypes evokes a similar role at the level of individual gene expression (Figure 3.1). However, the reality of the relationship between hormones and sex-biased expression bears surprisingly little resemblance to these intuitive expectations (Table 3.2). Though presence of HREs and expression bias are significantly associated, the directions of this association strongly depart from those of our conceptual model. This finding implies that current genomic and hormonal intuition is less accurate than we may like to admit. However, as models are only as good as their underlying assumptions, discordance between our results and model predictions should bring us to question these assumptions.

One possibly dubious assumption is that a gene's hormonal environment is an honest, sufficient, and possibly noise-free indicator of its actual sexual environment. While it is

generally true that testes and ovaries differentially produce androgens and estrogens, both circulate at appreciable concentrations in both sexes (Ketterson et al. 2005). The production of these hormones is also not limited to sex-specific gonadal tissues. Synthesis of steroid hormones in the brain can produce a local hormonal environment that may differ from that expected based on an individual's chromosomal sex (Schlinger et al. 2001). Additionally, local conversions of androgens to estrogens occur commonly. For example, male brain development is distally dependent upon circulating testosterone (an androgen) yet proximally determined in part by testosterone's local conversion into estradiol (an estrogen) by the enzyme aromatase (Davies and Wilkinson 2006). Describing this male-specific process as being androgen or estrogen-dependent is thus ambiguous. Therefore, how (or if) genes perceive their sexual environment in the midst of this hormonal uncertainty remains unclear.

Perhaps we are also assuming too large a role for hormones in determining sex-biased expression. This assumption, however, stands on a far more solid foundation. Another study also in *Mus* used multiple experimental procedures to investigate hormonal influences on adult sex-biased gene expression (van Nas et al. 2009). Specifically, these authors used 'four core genotype' mice to estimate the importance of hormones. These mice are genetically modified such that the normally Y-linked master sex determining gene (*Sry*) is removed and reinserted onto an autosome. Crosses performed with these mice allowed the authors to experimentally separate chromosomal sex from gonadal sex (i.e. by producing XY and XX males along with XX and XY females). Combined with gonadectomy and hormone-replacement, these procedures unambiguously showed that hormones have a substantial impact on sex-biased expression. Alongside this observation, one of our results in particular highlights an important departure from our intuitive framework on the nature of gene expression regulation.

Our simple scenario for the evolution of sex-biased expression posits a focal role for HREs acting in *cis*. Given hormones' known influence on sex-biased genes, it was thus odd to observe that so few of these genes contain hormone-sensitive regulatory elements (Table 3.2). The juxtaposition of these two findings suggests, in strong contrast to our scenario, that most sex-biased genes evolved regulation in *trans*. The unexpected prominence of *trans*-environments should therefore alter our thinking on sex-biased genes' evolutionary trajectories and how they may influence the process of conflict resolution.

Trans-effects: evidence and consequences

Male-biased genes were overrepresented among genes with expression-related molecular functions (Table 3.3). Sex-biased genes may thus play a disproportionate role in regulating the expression of other genes. This is especially relevant for other genes attempting to 'listen' for cues from their external sexual environment. *Trans*-acting sex-biased genes would be in a position to transfer their own sex-bias to downstream targets. Indeed, we found a weak correlation between the sex bias of a subset of the TFs in our dataset and the sex bias of downstream targets (Figure 3.2). However, many targets are regulated by multiple transcription factors, some with opposing sex-biases. Thus, genes must respond to several independent and possibly contradictory signals from both *trans* and hormonal environments to achieve sex-biased expression. Such inconsistent information may itself represent an additional form of antagonism.

In general, the genetic interests of a locus experiencing conflict need not be equivalent to the interests of other loci acting to resolve it (Burt and Trivers 1998; Wilkins and Haig 2002). Disagreement over the nature of such resolutions effectively gives rise to a new type of

antagonism. And though considerable evolution may occur in such a situation, little progress on conflict resolution may ultimately result at the original focal locus. The role of *trans*-environments inferred here suggests that such conflict born from conflict may be a common and important feature of sexually antagonistic selection. One example involves sex-biased transcription factors imposing their bias onto their targets. While in some circumstances this imposition might be favorable to both genes, in other cases their interests may diverge. Selection may then favor new, opposing regulatory elements acting in *cis*. This might give rise to antagonistic coevolution over time between sex-biased loci and their *trans*-acting regulators (perhaps analogous to conflicts between Imprinted and Imprinter genes, e.g. Burt and Trivers 1998). Similarly, interactions between multiple *trans*-regulators of a shared target gene may represent a new form of antagonism. Contradictory actions of *trans*-acting genes may spur antagonistic coevolution among themselves, with ultimately uncertain consequences for the expression of their common target. However, these *trans*-effects and the conflict they beget do not simply occur in a genomic vacuum. These interactions take place upon a chromosomal backdrop where different inheritance patterns also contribute to the nature of conflict resolution.

Differences in autosomal and X-chromosomal inheritance influence a wide range of biological phenomena. These phenomena range from speciation to kin-based cooperative behavior and dispersal, female mating preferences and male display traits, and the resolution of intralocus sexual conflict, among others (Haig 2000; Coyne and Orr 2004; Albert and Otto 2005; Frank and Crespi 2011). Regarding sexually antagonistic traits, X-linked genes are expected to favor female interests relative to autosomal genes (Frank and Crespi 2011; but see Patten and Haig 2009). This chromosomal context is especially relevant here in that we observe an overabundance of not only sex-biased genes, but *trans*-acting sex-biased genes on the X-

chromosome (Table 3.3). This observation suggests that not only should such chromosomal conflicts commonly occur within the genome, but also that conflict may spill over onto a larger set of target genes. Specifically, X-linked *trans*-acting genes may shape the outcome of conflict resolution experienced by autosomal genes and vice versa. Collectively, these *trans* interactions and the conflict they give rise to, along with the uncertainties of hormonal environments, should motivate us to rethink how sexual dimorphism and conflict resolution may evolve.

The timescale of conflict resolution

In order to understand the evolution of sex-biased expression at a locus, we must consider the effects of mutations both on the locus itself and at sites elsewhere in the genome. Our results indicate that these other sites in the genome are unexpectedly prevalent. These multiple *trans*-acting loci serve to increase the mutational target size for any focal gene experiencing sexual antagonism. For any given rate of mutation acting on this larger target, we therefore would expect sex-biased expression to evolve quickly, in accord with recent observations (Ranz et al. 2003; Jiang and Machado 2009). However, we must also consider the effect size of these mutations, acting both in *trans* and *cis*. Due to the number of interacting players within a gene's *trans* environment, the net effect of any particular *trans*-acting mutation may be relatively modest. Similarly, uncertainty within a gene's hormonal environment suggests that acquiring a *cis*-acting hormone response element may only moderately promote sex-biased expression on average. Together, these observations make contrasting predictions for the evolution of sex-biased expression and conflict resolution that differ from previous arguments (Stewart et al. 2010). On the one hand, a large mutational target allows sex-biased expression to evolve quickly

for a focal gene. But on the other hand, the generally weak sex-biasing effect of any particular *cis* or *trans* mutation would be unable to fully resolve antagonism at that gene. Therefore, while we would expect sexual dimorphism to evolve quickly, the resolution of antagonism it supposedly represents would actually evolve more slowly.

A consequence of this logic is that genome-wide sexual dimorphism should represent widespread unresolved sexual antagonism. However, studies that measure fitness are needed to directly assess the generality of this logic. One study that accomplished this Herculean task notably did not support this argument (Innocenti and Morrow 2010). Instead, these authors found that only 8.5% of sex-biased transcripts represent unresolved sexual antagonism in *Drosophila*. However, given the unusual influence of the Y-chromosome on genome-wide expression patterns in *Drosophila* (Lemos et al. 2008), and the sensitivity of sexually antagonistic selection to lineage-specific population genetic conditions (Connallon and Clark 2012), it remains to be seen whether this result is exceptional in a broader taxonomic context. In contrast, studies focusing on whole-organism sexually dimorphic phenotypes generally show dimorphism coexisting with estimates of ongoing sexually antagonistic selection across a diversity of animal taxa (Cox and Calsbeek 2009; Harano et al. 2010). This coexistence could be explained at a genetic level in part by the arguments proposed here.

Conclusion

Steroid sex hormones play a complex role in the production of sexual dimorphism and the process of conflict resolution. The intuitive notion of hormones operating simply through *cis*-acting response elements to bring about sex-biased gene expression has surprisingly little explanatory power genome-wide. Instead, these hormones must work both through and

alongside a previously underappreciated *trans*-regulatory environment. These *trans* influences both expand the target for mutations conferring sex-biased expression and increase the opportunity for new forms of antagonism. Collectively, this may result in the seemingly paradoxical combination of sexual dimorphism evolving quickly yet conflict resolution proceeding slowly. This scenario may thus explain the maintenance of sexually antagonistic fitness variation in the face of widespread sexual dimorphism that should otherwise eliminate it.

Acknowledgements

I would like to thank Manus Patten, Yoel Stuart, and Andrew Richardson for their constructive criticism on previous versions of this chapter.

References

- Agresti, A. 2007. An introduction to categorical data analysis, 2nd ed. John Wiley & Sons, New York, NY.
- Albert, A. Y. K. and S. P. Otto. 2005. Sexual selection can resolve sex-linked sexual antagonism. *Science* 310: 119-121.
- Arnold, A. P. 2009. The organizational-activational hypothesis as the foundation for a unified theory of sexual differentiation of all mammalian tissues. *Hormones and Behavior* 55: 570-578.
- Bonduriansky, R. and S. F. Chenoweth. 2009. Intralocus sexual conflict. *Trends in Ecology & Evolution* 24: 280-288.
- Bourdeau, V., J. Deschenes, R. Metivier, Y. Nagai, D. Nguyen, N. Bretschneider, F. Gannon, J. H. White and S. Mader. 2004. Genome-wide identification of high-affinity estrogen response elements in human and mouse. *Molecular Endocrinology* 18: 1411-1427.
- Burt, A. and R. Trivers. 1998. Genetic conflicts in genomic imprinting. *Proceedings of the Royal Society B: Biological Sciences* 265: 2393-2397.
- Claessens, F., G. Verrijdt, E. Schoenmakers, A. Haelens, B. Peeters, G. Verhoeven and W. Rombauts. 2001. Selective DNA binding by the androgen receptor as a mechanism for hormone-specific gene regulation. *Journal of Steroid Biochemistry and Molecular Biology* 76: 23-30.
- Connallon, T. and A. G. Clark. 2012. A general population genetic framework for antagonistic selection that accounts for demography and recurrent mutation. *Genetics* 190: 1477-1489.
- Couse, J. F., S. C. Hewitt, D. O. Bunch, M. Sar, V. R. Walker, B. J. Davis and K. S. Korach. 1999. Postnatal sex reversal of the ovaries in mice lacking estrogen receptors alpha and beta. *Science* 286: 2328-2331.
- Cox, R. M. and R. Calsbeek. 2009. Sexually antagonistic selection, sexual dimorphism, and the resolution of intralocus sexual conflict. *The American Naturalist* 173: 176-187.
- Coyne, J. A., E. H. Kay and S. Pruett-Jones. 2008. The genetic basis of sexual dimorphism in birds. *Evolution* 62: 214-219.
- Coyne, J. A. and H. A. Orr. 2004. Speciation. Sinauer Associates, Inc., Sunderland, MA.
- Davies, W. and L. S. Wilkinson. 2006. It is not all hormones: alternative explanations for sexual differentiation of the brain. *Brain Research* 1126: 36-45.
- Frank, S. A. and B. J. Crespi. 2011. Pathology from evolutionary conflict, with a theory of X chromosome versus autosome conflict over sexually antagonistic traits. *Proceedings of the National Academy of Sciences* 108: 10886-10893.

- Fry, J. D. 2010. The genomic location of sexually antagonistic variation: some cautionary comments. *Evolution* 64: 1510-1516.
- Galani, A., S. Kitsiou-Tzeli, C. Sofokleous, E. Kanavakis and A. Kalpini-Mavrou. 2008. Androgen insensitivity syndrome: clinical features and molecular defects. *Hormones-International Journal of Endocrinology and Metabolism* 7: 217-229.
- Gregg, C., J. W. Zhang, B. Weissbourd, S. J. Luo, G. P. Schroth, D. Haig and C. Dulac. 2010. High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science* 329: 643-648.
- Gruber, C. J., D. M. Gruber, I. M. L. Gruber, F. Wieser and J. C. Huber. 2004. Anatomy of the estrogen response element. *Trends in Endocrinology and Metabolism* 15: 73-78.
- Haig, D. 2000. Genomic imprinting, sex-biased dispersal, and social behavior. *Annals of the New York Academy of Sciences* 907: 149-163.
- Harano, T., K. Okada, S. Nakayama, T. Miyatake and D. J. Hosken. 2010. Intralocus sexual conflict unresolved by sex-limited trait expression. *Current Biology* 20: 2036-2039.
- Innocenti, P. and E. H. Morrow. 2010. The sexually antagonistic genes of *Drosophila melanogaster*. *PLoS Biology* 8: e1000335.
- Jiang, Z. F. and C. A. Machado. 2009. Evolution of sex-dependent gene expression in three recently diverged species of *Drosophila*. *Genetics* 183: 1175-1185.
- Ketterson, E. D., V. Nolan and M. Sandell. 2005. Testosterone in females: mediator of adaptive traits, constraint on sexual dimorphism, or both? *The American Naturalist* 166: S85-S98.
- Korach, K. S. 1994. Insights from the study of animals lacking functional estrogen-receptor. *Science* 266: 1524-1527.
- Lande, R. 1980. Sexual dimorphism, sexual selection, and adaptation in polygenic characters. *Evolution* 34: 292-305.
- Lemos, B., L. O. Araripe and D. L. Hartl. 2008. Polymorphic Y chromosomes harbor cryptic variation with manifold functional consequences. *Science* 319: 91-93.
- Mank, J. E. 2009. Sex chromosomes and the evolution of sexual dimorphism: lessons from the genome. *The American Naturalist* 173: 141-150.
- Moehren, U., S. Denayer, M. Podvinec, G. Verrijd and F. Claessens. 2008. Identification of androgen-selective androgen-response elements in the human aquaporin-5 and Rad9 genes. *Biochemical Journal* 411: 679-686.
- Notini, A. J., R. A. Davey, J. F. McManus, K. L. Bate and J. D. Zajac. 2005. Genomic actions of the androgen receptor are required for normal male sexual differentiation in a mouse model. *Journal of Molecular Endocrinology* 35: 547-555.

- Parisi, M., R. Nuttall, D. Naiman, G. Bouffard, J. Malley, J. Andrews, S. Eastman and B. Oliver. 2003. Paucity of genes on the *Drosophila* X chromosome showing male-biased expression. *Science* 299: 697-700.
- Patten, M. M. and D. Haig. 2009. Maintenance or loss of genetic variation under sexual and parental antagonism at a sex-linked locus. *Evolution* 63: 2888-2895.
- Patten, M. M., D. Haig and F. Úbeda. 2010. Fitness variation due to sexual antagonism and linkage disequilibrium. *Evolution* 64: 3638-3642.
- Poissant, J., A. J. Wilson and D. W. Coltman. 2010. Sex-specific genetic variance and the evolution of sexual dimorphism: a systematic review of cross-sex genetic correlations. *Evolution* 64: 97-107.
- R Development Core Team. 2011. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at <http://www.R-project.org>.
- Ranz, J. M., C. I. Castillo-Davis, C. D. Meiklejohn and D. L. Hartl. 2003. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science* 300: 1742-1745.
- Rhen, T. 2000. Sex-limited mutations and the evolution of sexual dimorphism. *Evolution* 54: 37-43.
- Rice, W. R. 1984. Sex-chromosomes and the evolution of sexual dimorphism. *Evolution* 38: 735-742.
- Rice, W. R. and A. K. Chippindale. 2001. Intersexual ontogenetic conflict. *Journal of Evolutionary Biology* 14: 685-693.
- Schlinger, B. A., K. K. Soma and S. E. London. 2001. Neurosteroids and brain sexual differentiation. *Trends in Neuroscience* 24: 429-431.
- Stewart, A. D., A. Pischedda and W. R. Rice. 2010. Resolving intralocus sexual conflict: genetic mechanisms and time frame. *Journal of Heredity* 101: S94-S99.
- van Doorn, G. S. 2009. Intralocus sexual conflict. *Annals of the New York Academy of Sciences* 1168: 52-71.
- van Nas, A., D. GuhaThakurta, S. S. Wang, N. Yehya, S. Horvath, B. Zhang, L. Ingram-Drake, G. Chaudhuri, E. E. Schadt, T. A. Drake, A. P. Arnold and A. J. Lusis. 2009. Elucidating the role of gonadal hormones in sexually dimorphic gene coexpression networks. *Endocrinology* 150: 1235-1249.
- Verrijdt, G., A. Haelens and F. Claessens. 2003. Selective DNA recognition by the androgen receptor as a mechanism for hormone-specific regulation of gene expression. *Molecular Genetics and Metabolism* 78: 175-185.

- Wang, P. J., J. R. McCarrey, F. Yang and D. C. Page. 2001. An abundance of X-linked genes expressed in spermatogonia. *Nature Genetics* 27: 422-426.
- Wilkins, J. F. and D. Haig. 2002. Parental modifiers, antisense transcripts and loss of imprinting. *Proceedings of the Royal Society B: Biological Sciences* 269: 1841-1846.
- Zhang, H. M., H. Chen, W. Liu, H. Liu, J. Gong, H. L. Wang and A. Y. Guo. 2012. AnimalTFDB: a comprehensive animal transcription factor database. *Nucleic Acids Research* 40: D144-D149.

CHAPTER 4

Heterochiasmy and genomic imprinting: evolutionary and molecular perspectives

Abstract

The observation of quantitative differences in the rate of recombination between the sexes, a phenomenon known as heterochiasmy, has been a difficult phenomenon to explain for decades. More recent population genetic modeling has shown that sex-differences in epistasis predominantly acting in the haploid stage of the life cycle can promote the evolution of heterochiasmy. In mammals, imprinted genes and their epistatic partners are one set of genes that can experience haploid selection despite acting in the diploid phase of the life cycle. Previous empirical verification of these theoretical predictions has been limited both in terms of taxonomic scope and the resolution of data on imprinting and recombination. Using new high-resolution data from the lab mouse and more sensitive spatially-explicit analytical approaches, we reevaluated these theoretical predictions while adjusting for a variety of molecular aspects of the genome. Coincident with these predictions, we found a modest association between imprinting and the magnitude of sex-differences in recombination rates. However, substantial variation in recombination remained unexplained, due in part to the minimal contribution of the other genomic variables. In light of these results, we discuss other genomic factors possibly serving as the proximate basis for heterochiasmy and how such factors might influence its evolution. Similarly, we discuss how heterochiasmy specifically and variation in recombination more generally can intertwine the evolutionary futures of genes experiencing previously independent selective pressures.

Introduction

The phenomenon of two separate sexes sharing a common genome poses interesting evolutionary and genetical problems. Ecological differences between the sexes and numerous reproductive asymmetries give rise to sex-specific selective pressures across a wide variety of taxa. A consequence of these divergent pressures is our expectation of a *negative* correlation for fitness between males and females across a variety of phenotypes. On the other hand, as a result of a single shared genome, we also expect and commonly measure a strongly *positive* genetic correlation for phenotypes shared between the sexes (Bonduriansky and Chenoweth 2009; van Doorn 2009). The juxtaposition of these two observations seems to represent a substantial impediment to sex-specific adaptation (Lande 1980). However, a cursory examination of rampant phenotypic sex differences in nature would suggest that evolution has commonly and often quickly overcome such impediments. Nevertheless, sex differences observed in the most basic and fundamental cellular processes underlying reproduction should still pique our interest. Meiosis is one such process.

Meiosis is an ancient and complex cellular phenomenon determining fertility across a large swath of sexual taxa. Like any large-scale cellular activity, it is composed of a number of sub-processes, the results or aspects of which can vary dramatically across the sexes. The production of different gametes, anisogamy, is probably the most obvious and well-known result of meiosis that characterizes the sexes, and prior theoretical work has shed considerable light on the features of its evolution (Parker et al. 1972; Cosmides and Tooby 1981). Similarly, the asymmetric nature of female meiosis, in which only one of the four possible products becomes a functional gamete, is fairly well understood, with interesting implications for meiotic drive systems having also been worked out (Pardo-Manuel de Villena and Sapienza 2001; Haig 2010;

Brandvain and Coop 2012). However, our understanding of many other dimorphic aspects of meiosis is surprisingly lacking due in no small part to a basic lack of information across taxa.

One noteworthy example would be sexual dimorphism in the timing of meiosis. In mammals, females begin meiosis quite early in life as a developing fetus, yet its progression is arrested relatively early during prophase I. These arrested oocytes wait for potentially several decades (in humans) before ovulation and then complete meiosis II only upon successful fertilization. In contrast, male meiosis begins shortly before reproductive maturity and thereafter proceeds in regular waves for the remainder of adult life (Morelli and Cohen 2005). Although this pattern is thought to be generally true for all mammals, this generalization is based upon an examination of relatively few species. Worse still, sampling more broadly among other vertebrates seems to be lacking, providing no evidence for whether this phenomenon is actually restricted to mammals. This general lack of data in part comes from the difficulty of studying meiosis *in utero* and from the field's expected bias towards work on a small set of model organisms. Given this informational vacuum, it is perhaps unsurprising that no theoretical explanation seems to exist for why such dimorphism would evolve in the first place. In contrast, we have a slightly better grasp on another similarly underappreciated aspect of meiosis, the difference in recombination between the sexes.

Sex differences in the rate of recombination are commonly observed in nature and collectively fall into one of two mechanistically and evolutionarily distinct categories. One such category is achiasmy, in which one of the two sexes experiences no recombination at all. In certain taxa, achiasmy is a trivially necessary consequence of their sex determination mechanisms. For example, males of haplodiploid species have no homologous set of chromosomes needed for crossing over, necessarily limiting recombination solely to diploid

females. In many other normally diploid taxa, however, achiasmy is almost exclusively a feature of the heterogametic sex (Burt et al. 1991). The second category, heterochiasmy, consists of quantitative differences in the rates of recombination between the sexes. Although it may seem that achiasmy would simply be an extreme case of heterochiasmy, a variety of factors differentiate the two phenomena.

Mechanistic and taxonomic considerations partially differentiate achiasmy from heterochiasmy. For example, in heterochiasmate taxa, homologous chromosomes that fail to form a single crossover often lead to the formation of generally deleterious aneuploid gametes (Hassold and Hunt 2001). However, similar and widespread patterns of aneuploidy are clearly not seen in achiasmate taxa like *Drosophila*, despite a complete lack of crossing over among all pairs of homologous chromosomes in males (Hartl et al. 2008). Regarding taxonomic patterns, they are also inconsistent between achiasmy and heterochiasmy. Specifically, strong associations between achiasmy and sex chromosomes are generally not found with heterochiasmate taxa. Heterochiasmate taxa typically demonstrate female-biased recombination rates, though considerable lineage specificity in these patterns is evident (Burt et al. 1991; Lenormand and Dutheil 2005; Haig 2010).

Evolutionary explanations for the two phenomena also differ considerably. The generally accepted explanation for autosomal achiasmy is that it is a pleiotropic consequence of recombination suppression between sex chromosomes in the heterogametic sex (Haldane 1922). However, such notions fail to explain subtle quantitative sex differences, occasional examples of higher recombination in the heterogametic sex, heterochiasmy in hermaphrodites, and heterochiasmy in taxa lacking sex chromosomes (Lenormand and Dutheil 2005). More recent attempts at explaining the evolution of heterochiasmy have therefore focused on a variety of

other biological factors ranging from meiotic drive and epistasis to the vagaries of sexual selection. In this chapter, we mostly restrict our attention to a study that focused on epistasis as the most general cause of linkage disequilibrium typically needed for the evolution of recombination.

Specifically, we focus on a recent population genetic model investigating the conditions under which natural selection would favor sex-specific changes in the rate of recombination (Lenormand 2003). This model investigated the strength of selection acting on a locus capable of modifying the recombination rate between two other loci subject to viability selection acting in both the diploid and haploid phases of the life cycle. In the diploid phase, results of the model suggest that highly specialized and unusual conditions are required to favor the evolution of heterochiasmy. This result also has negative implications for prior ideas on why heterochiasmy might evolve. In particular, Trivers theorized that when there is differential variance in reproductive success, selection would favor reduced recombination in the sex with greater variance (typically males) (Trivers 1988). Specifically, this verbal model suggests that since relatively few males achieve any reproductive success, selection would favor maintaining the associations between alleles contributing to that success. While such reasoning has considerable intuitive appeal for such a difficult evolutionary problem, empirical evidence on the matter is mixed (Burt et al. 1991; Mank 2009), which could very well be a consequence of diploid selection's lack of efficacy in favoring heterochiasmy. In contrast, Lenormand's model results suggest that sex differences in epistasis in the haploid phase are particularly conducive to the evolution of heterochiasmy.

As many animals spend little to no time subject to selection in the haploid phase of their life cycles, it may seem at first that these theoretical results would have little relevance for such

species. But among mammals at least, genes subject to genomic imprinting are special in their ability to experience haploid selection despite being expressed in the diploid phase of the life cycle. On the other hand, imprinted genes are typically thought to represent a small minority of all genes in the genome, seemingly limiting their role in the evolution of heterochiasmy.

However, epistatic interactions between imprinted and far more common unimprinted genes might drastically increase the fraction of the genome that could be affected by this kind of quasi-haploid selection (Joseph and Kirkpatrick 2004). Indeed, in a recent reanalysis of 97 phenotypes of outbred mice, widespread parent-of-origin effects were identified that had previously escaped detection (Mott et al. 2014). Further analyses by the authors suggested that precisely this form of epistasis between imprinted and unimprinted genes could be a likely cause of such common parent-of-origin effects. Together, these theoretical predictions and empirical observations suggest that imprinting could have more pervasive effects on the general recombinational landscape of the genome than has previously been appreciated.

Unfortunately, these predictions have received relatively little empirical attention over time, due in part to various difficulties with the availability of relevant data. For example, estimating recombination rates of sufficient resolution can be time-consuming and expensive and thus is typically limited to humans and model taxa in which linkage map construction is a feasible activity. Data do exist for many other species though, based on counts of chiasmata or MLH1 foci (a marker of late recombination nodules). However, such cytological data are also difficult to gather and are relevant more for questions on genome-wide recombination rates across taxa than on finer-scale recombination rates within chromosomes (Dumont and Payseur 2011). Similarly, genomic imprinting has not been a trivial phenomenon to measure. Prior to the advent of high-throughput estimation of allele-specific expression patterns, most imprinted

genes were discovered individually through laborious experimental approaches and subsequent validation. Slow growth in the number of known imprinted genes also created considerable uncertainty over time in any estimate of their overall abundance in the genome. And again, this work took place mostly in mouse models and in humans, generally limiting the extent of any cross-species comparisons.

Difficulties also exist regarding the measurability of other relevant data. To test the predictions of the model described above, ideally we would also like to have solid genome-wide estimates of epistasis between imprinted and unimprinted genes. However, estimating epistasis can be a substantial and difficult undertaking. Measurements of statistical epistasis are made in the context of a population and generally are limited to only the most tractable of model systems, particularly in the context of epistasis for fitness (Corbett-Detig et al. 2013). Functional epistasis, on the other hand, is independent of a population context and instead focuses on the molecular interactions between proteins or other products of different genes. Construction of such resources at genome-wide scales is an ongoing task for model organisms generally and the lab mouse specifically (Guan et al. 2008; Iossifov et al. 2009; Li et al. 2010). These databases are typically constructed using text-mining of the literature and machine learning approaches that probabilistically assign relationships between gene products. However, it is currently unclear whether such indirect approaches would agree with more direct measurements of gene interaction, such as those derived from yeast two-hybrid screening. Worse still, information on this kind of functional epistasis may be uninformative on statistical epistasis that would be more evolutionarily relevant (Phillips 2008). Finally, even if we had our ideal dataset on statistical epistasis between imprinted and unimprinted genes, this contemporary data would still only represent an uncertain estimate of the history of epistasis experienced by an organism.

Epistasis is also not the only factor potentially influencing recombination rates. Various molecular features of DNA sequence itself and the higher-order chromatin configuration of chromosomes can or are thought to affect the process of crossing over. Such features range from GC content to local densities of repetitive elements to the proximity of the centromere, among many others (Lercher and Hurst 2003; Jensen-Seaman et al. 2004; Sandovici et al. 2006; Dumont et al. 2011). However, causal connections between any of these variables and recombination rates do not preclude the influence of epistasis on such rates. As explanatory factors, these proximate features of the chromosomal environment would instead complement ultimate aspects of the epistatic environment's prior influence on recombination through the process of natural selection (Mayr 1961). Nevertheless, such variables can therefore serve as important covariates to consider in a new assessment of the connections between epistasis, genomic imprinting, and heterochiasmy.

Prior empirical research on these associations is essentially limited to two studies in humans making use of both a subset of currently known imprinted genes and fairly coarse measures of sex-specific recombination rates (Lercher and Hurst 2003; Necsulea et al. 2009). Given more recent advances and an expanded set of data that can be brought to bear on existing theory, we feel there is a good opportunity at the moment to reevaluate these associations more effectively. Specifically, while considering a variety of covariates derived from a recent build of the mouse genome, we intend to utilize new estimates of genomic imprinting to test theoretical predictions regarding these genes' positive influence on the magnitude of sex differences in local recombination rates. Additionally, we will test predictions regarding the nature of imprinting (maternal or paternal silencing) and the direction of sex differences in recombination. Finally, we will also describe and utilize a novel implementation of spatial statistics in these analyses to

help address a problem of non-independence among genes that may be potentially prevalent in studies conducted at whole-genome scales.

Methods

Recombination rate data and estimation

In order to address the theoretical predictions set out above, we first acquired high-quality data on recombination rates estimated separately in each sex. A recent study using a large-scale cross of heterogeneous stock strains of *Mus* recently improved on prior estimates of the linkage map in this species (Cox et al. 2009). Using data from over 3,500 meioses, this study produced male, female, and sex-averaged linkage maps with considerable sub-centiMorgan resolution. Following this study, we then re-estimated local broad-scale recombination rates (e.g. cM/Mb) in all three linkage maps with a sliding window approach. The genetic and physical positions of mapping SNPs within 5 Mb windows were used to calculate point estimates of the local recombination rate. Each rate estimate was then associated with a physical location at the midpoint of each window. A 0.25 Mb step was then taken from the beginning of the prior window and this process was repeated across the entirety of all autosomes.

As the analysis here ultimately resides at the level of individual genes, we then used nearest-neighbor interpolation to assign rate information to gene models specified in the UCSC mm9 genome knownGene table (Fujita et al. 2011). Specifically, the physical midpoint of each gene was assigned the recombination rate of the nearest sliding window midpoint. Such an approach may initially seem too coarse given the single base pair resolution of features within the thoroughly annotated mm9 genome. However, the lower relative resolution of these linkage

maps and the focus on broad-scale recombination rates would suggest little benefit to implementing more sophisticated interpolation methods whose results should be largely coincident with the approach used here.

Using this approach, we then had information available on female, male, and sex-averaged recombination rates for all the genes contained within the mapped region of the original study (approximately 98% of the entire *Mus* genome) (Cox et al. 2009). However, given the proximity of genes to one another along their chromosomes and the inherent nature of sliding window-based estimation, this data, and any other basic summary statistic constructed from it, suffers from a potentially high degree of non-independence that could bias downstream analyses. Specifically, recombination rate estimates for these genes should be positively spatially autocorrelated to some extent. Neighboring windows in the sliding window analysis will share some fraction of the same mapping SNPs depending on the number of steps between each window and the densities of SNPs within each of these windows. Therefore, nearby genes might have more similar estimates of their recombination rates than would be expected by chance. Thankfully, approaches exist both to estimate this degree of autocorrelation and eliminate some to all of its biasing effects on subsequent analyses.

Specifically, we used Moran's I statistic to estimate the degree of spatial autocorrelation in recombination rates among these genes (Moran 1948; Moran 1950; Cliff and Ord 1973; 1981). As described in Chapter 2, Moran's I is a global measure of autocorrelation of any single variable whose observations exist at fixed positions in space, shown in the equation below.

$$I = \frac{n \sum_{i=1}^n \sum_{j=1, j \neq i}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\left(\sum_{i=1}^n \sum_{j=1, j \neq i}^n w_{ij} \right) \left(\sum_{i=1}^n (x_i - \bar{x})^2 \right)}$$

In our case, these fixed positions are the known physical locations of genes along their chromosomes and the variable of interest can be some composite measure of recombination rate within or between the sexes. The main variable we use here to capture the magnitude of heterochiasmy is the absolute value of the difference between female and male recombination rates. However, in order to measure possible spatial autocorrelation for this variable, we must first consider how spatial dependencies might arise in the first place and how this can be represented mathematically.

Representing spatial relations between genes mathematically requires constructing a generalized spatial weight matrix that represents the pairwise relationship of all the genes in our sample. As the biology of recombination and the methodological details of our sliding window approach would imply no spatial relationships between genes on different chromosomes, we constructed separate weight matrices and performed separate analyses for each of the 19 autosomes of the *Mus* genome.

Within each chromosome, we expect potential patterns of autocorrelation among genes to decay linearly with distance. This linear decay would be a consequence of the movement of windows along each chromosome. Specifically, the 5 Mb window used to estimate the local recombination rate moves along the chromosome in regular 0.25 Mb steps. Relative to a given fixed window, subsequent windows would therefore overlap with the original window to a decreasing extent with each step taken. At a distance of 20 steps between window midpoints, recombination rate estimates would be based on entirely non-overlapping sets of mapping SNPs. We can then expect that estimates for genes at this or greater distances would be entirely spatially independent, whereas genes located less than the distance of a single window step would likely receive estimates based on the exact same mapping SNP data. For this latter

example, some uncertainty exists due to the interpolation procedure described above. Two genes located within 0.25 Mb may nevertheless be assigned to different adjacent window midpoints depending on the relative distances from each gene to their nearest window midpoint. To account for this aspect of the interpolation process, we define the parameter d_{ij} to represent the floored distance between genes i and j measured in steps. Taken together, we can then represent the spatial decay of autocorrelation between genes i and j (where $i \neq j$ and $w_{ii} = 0$) using the following expression.

$$w_{ij} = 1 - d_{ij}/20$$

For each autosome, we then calculated a standardized form of Moran's I bounded between -1 and $+1$, where values closer to $+1$ imply strong positive spatial autocorrelation among genes' estimated recombination rates. Although the recombination data available here represents a very large proportion of autosomal genes, we restricted this autocorrelation analysis to the subset of genes for which we also estimated imprinted gene expression (12,625 genes whose analysis is described in a subsequent section). Statistical significance of Moran's I was then assessed by comparing Z -scores to the standard normal distribution, as I is asymptotically normally distributed (Cliff and Ord 1973; 1981). We used a one-tailed alternative hypothesis for these tests, as we have no biological or methodological expectation of observing any possibility of negative spatial autocorrelation. Therefore, chromosomes with Z -scores greater than 1.65 (representing a standard p -value of 0.05) were considered significantly autocorrelated. However, given the magnitudes of the Z -scores shown in Table 4.1 below, our specific threshold of significance is relatively unimportant here.

Table 4.1. Autocorrelation results for the magnitude of sex-differences in recombination.

| Chromosome | Sample size (<i>n</i>) | Standardized <i>I</i> | Z-score |
|-------------------|-------------------------------|------------------------------|----------------|
| 1 | 763 | 0.746 | 82.6 |
| 2 | 1028 | 0.617 | 60.0 |
| 3 | 623 | 0.639 | 42.5 |
| 4 | 829 | 0.831 | 132.0 |
| 5 | 861 | 0.798 | 116.6 |
| 6 | 662 | 0.734 | 71.1 |
| 7 | 967 | 0.657 | 85.7 |
| 8 | 708 | 0.679 | 78.2 |
| 9 | 751 | 0.821 | 133.8 |
| 10 | 659 | 0.669 | 81.1 |
| 11 | 1042 | 0.873 | 198.6 |
| 12 | 489 | 0.658 | 45.0 |
| 13 | 466 | 0.629 | 36.4 |
| 14 | 475 | 0.666 | 56.2 |
| 15 | 520 | 0.871 | 153.8 |
| 16 | 393 | 0.744 | 41.5 |
| 17 | 618 | 0.697 | 86.2 |
| 18 | 352 | 0.718 | 55.6 |
| 19 | 419 | 0.610 | 61.3 |

As expected, the magnitude of positive spatial autocorrelation in recombination rates is quite high for all autosomes. Were this to go uncorrected in subsequent analyses, such results would likely be considerably biased. Specifically, autocorrelated variables have a known influence on general linear models. Although the coefficients of such models are accurately estimated in the presence of autocorrelation, standard errors of those estimates tend to be underestimated (Martin 1974). Such underestimation would inflate the significance of model coefficients, thus leading to an increased occurrence of false positives. Thankfully, approaches are available to address this type of problematic autocorrelation and decrease its influence on statistical inference.

One relatively simple approach that deals with autocorrelation is spatial differencing. This approach involves the creation of a new variable in which the weighted effects of neighboring observations are subtracted out from the original observations (Martin 1974). If the

original observations are elements of the vector y , then the spatially differenced version of that variable, y' , can be calculated using the following expression:

$$\bar{y}' = \bar{y} - p\mathbf{W}\bar{y}$$

In this expression, the matrix \mathbf{W} represents the row-standardized version of the weight matrix described previously that captures the linear decline of spatial relationships with increasing distance between genes. The scalar p represents the autocorrelation parameter, here taken to be the standardized form of Moran's I from Table 4.1. As this parameter varies across chromosomes, spatial differencing was performed separately for each. Finally, similar analyses were performed on other recombination rate data. Specifically, spatially differenced versions of the average and a signed sex difference (female – male) in recombination rate were calculated as well.

Imprinted gene expression estimation

To assess the connection between imprinted genes and various measures of recombination rate variation, we gathered data from a recent study of the *Mus* transcriptome making use of a reciprocal cross between two relatively divergent strains (Gregg et al. 2010). Specifically, this study performed a cross between inbred males and females of the C57BL/6J strain of *Mus musculus domesticus* and the CAST/EiJ strain of *Mus musculus castaneus*. The natural populations from which these inbred strains were derived are estimated to have diverged approximately 500,000 years ago (Geraldes et al. 2008), making them particularly useful for estimating allele-specific expression here due to the large number of fixed differences between them. Additionally, this study performed transcriptome sequencing in both parent populations,

allowing SNPs measured in the F1 population to be experimentally validated as real variants inherited from each parent. Finally, this study looked at the transcriptomes of two separate regions of the adult *Mus* brain, the medial prefrontal cortex and preoptic area of the hypothalamus. Given their involvement in a variety of social behaviors and in the regulation of socially-relevant physiological processes (Gregg et al. 2010), these regions represent excellent expected locations for imprinted gene expression.

In order to estimate the direction and degree of imprinted expression for the purposes of this study, we first obtained a database of previously aligned RNA sequencing data from the first author of the previous study (Gregg et al. 2010). This database represented data that had already gone through a variety of quality control procedures, Novoalign-based alignment to the mm9 genome (Novocraft), and custom script-based SNP calling and validation by comparison with similarly processed parental sequences. The data were thus counts of reads containing a SNP carrying information on its parental origin for both male and female F1 adults resulting from the original ($\text{♀ CAST/EiJ} \times \text{♂ C57BL/6J}$) and reciprocal crosses ($\text{♀ C57BL/6J} \times \text{♂ CAST/EiJ}$). The data was then further processed to remove SNPs within a distance of less than the length of a typical sequencing read (about 32 base pairs) from another SNP, as measures of expression at such SNPs would not be independent.

Using such data, we then used a slightly different approach than the original authors to estimate differential expression between maternally and paternally derived alleles. As the focus of this study resides at the level of the gene and not at individual SNPs, we first summed the counts of SNP-containing reads across all SNPs for a given gene within each experimental condition. We then pooled male and female gene-level read counts together across both directions of the cross to ultimately derive single numbers representing counts of the maternally

and paternally derived alleles among the entire F1 population. For each gene, we then performed an exact binomial test on these counts, testing for significant deviations from a 1:1 expression pattern of the maternally and paternally derived alleles that one would expect under the null hypothesis of no imprinting. These procedures were performed independently for each brain region, resulting in up to two estimates of imprinting for each gene expressed. The resulting p -values from these exact tests were then used to calculate q -values in order to address the false discovery rate (Storey and Tibshirani 2003).

Significantly biased genes in one or both brain regions with a q -value less than 0.01 were then provisionally considered imprinted. However, as in the estimation of sex-biased expression discussed in previous chapters, subtle parental biases in highly expressed genes can nevertheless appear highly significant. From a strictly biological perspective, such subtle biases are unlikely to be a result of real genomic imprinting. In previous theoretical work, the “loudest voice prevails” principle suggests that the evolutionary end-point of an imprinted gene should essentially be an all-or-none expression level of its alleles depending on their parental origin (Haig 1996). Although such models do formally treat the evolution of imprinting as a continuous phenomenon, partial imprinting is only expected to be a transient state as the system moves towards this long-term equilibrium (Mochizuki et al. 1996). However, any empirical measurement of a real system could conceivably be taken at a slice of time in which the system has yet to reach such a point. This would suggest that the partial imprinting we view here could plausibly be real examples of genomic imprinting.

On the other hand, given that imprinting (and gene expression more generally) is inherently a cellular phenomenon, experimental error occurring at higher levels of organization may represent more likely causes of partial biases in expression. For example, the brain is

known to be a considerably heterogeneous organ in terms of the number and spatial arrangement of its cell-types (Kamme et al. 2003). A signal of complete imprinting of a gene in one cell-type could therefore be dampened by biallelic expression of the same gene in physiologically distinct cells located nearby. Combined with practical requirements of the RNA extraction process, biases introduced randomly during amplification, and subtle variation in read depth across samples during the sequencing process, this suggests that subtle parental biases are likely technical rather than biological in origin.

To address this degree of uncertainty, we used a series of bias thresholds in order to classify genes as imprinted or not. Our most liberal threshold consists only of the use of q -values to determine imprinting. The intermediate threshold considers both q -values and a larger than 20% difference in expression to diagnose imprinting. Finally, our more conservative and biologically most plausible threshold diagnoses genes as imprinted if $q < 0.01$ and they exhibit at least a two-fold difference in expression between the maternally and paternally derived alleles. Given the theoretical expectation of all-or-none expression for the alleles at an imprinted gene, we feel most confident about this final threshold. Additionally, we feel that this is a good compromise between higher, potentially more biologically realistic expression thresholds and the increased rate of false negatives that would follow with their use.

Finally, we also used entirely separate estimates of imprinting to address this uncertainty in a qualitatively different manner. Specifically, we made use of lists of previously discovered imprinted genes from curated databases (Mousebook, Geneimprint). These databases are manually curated lists of imprinted genes gathered from the literature along with the direction of their parental bias. In subsequent analyses, these genes and genes directly estimated here received a binary classification of 1 if imprinted and 0 if unimprinted. For a subset of analyses

that directly make use of the direction of allelic silencing, imprinted genes were qualitatively classified as maternally or paternally biased.

Estimation of other genomic variables

Although not explicitly involved in the theoretical predictions tested here, a number of other genomic variables are known to influence or have previously been correlated with recombination rates and/or genomic imprinting (Lercher and Hurst 2003; Jensen-Seaman et al. 2004; Sandovici et al. 2006; Dumont et al. 2011). These variables for the most part largely capture molecular aspects of the primary DNA sequence landscape. Primarily, these variables will simply serve as covariates in the analyses performed here due to the uncertainty of their causal association with recombination and their complementary relationship with the more evolutionarily relevant main variables described above. Nevertheless, their inclusion in subsequent linear models (described in the next section) may provide some reduction of residual error variation and thus are of some value here.

The majority of the genomic variables described below are derived directly from annotations of the mm9 genome in the UCSC Genome Browser. The first such variable is the distance in base pairs from each gene's canonical transcription start site to the nearest CpG island. Several density variables of repetitive elements were also collected based on previous RepeatMasker annotations of the mm9 genome (RepeatMasker Home Page). Specifically, the density of LINEs, SINEs, LTR elements, and simple repeats (such as microsatellites) were calculated in the vicinity of each gene investigated here. Some previous evidence has suggested that these repetitive elements could mediate attachment of chromatin to the synaptonemal

complex during prophase I of meiosis and thus the location of recombination events, although other evidence suggests a less-clear cut relationship (Heng et al. 1996). A measure of coding sequence compactness within each gene was also calculated based on previous suggestions of a connection between intron lengths and genomic imprinting (McVean et al. 1996). Additionally, the proportional distance of each gene to the centromere was calculated based on the physical midpoints of each gene. The *Mus* genome consists entirely of acrocentric chromosomes with largely heterochromatic sequence on the short arm of each (Chinwalla et al. 2002). Therefore, our calculations for genes on the euchromatic long-arms of these chromosomes simply ignore the presence of such short arms in the total length of each chromosome. Finally, this is an especially important covariate to include here as it is known that sex differences in recombination rate tend to become more male-biased towards the distal telomere (Petkov et al. 2007; Paigen and Petkov 2010), providing a qualitatively different form of spatial autocorrelation than those previously taken into account. Additionally, prior theoretical work has highlighted distance from the centromere as an important factor for recombination-modifying meiotic drive elements attempting to bias their transmission to the single functional gamete produced in female meiosis (Haig 2010; Brandvain and Coop 2012).

The last of our genomic covariates is the local GC content in the vicinity of each gene, since previous studies have implicated connections between various aspects of recombination and GC% (Lercher and Hurst 2003; Jensen-Seaman et al. 2004; Smagulova et al. 2011). As the structure of isochores is typically on the scale of hundreds of kilobases to megabases (Eyre-Walker and Hurst 2001), we used a sliding window approach to estimate GC% within 200 kb windows offset by 10 kb steps. Again, nearest neighbor interpolation approaches were used to assign GC content values centered on window midpoints to their nearest genes. Although once

again we might expect this sliding window approach to introduce potentially strong positive autocorrelation among sites for the estimates of GC%, this should be less of a concern here than it was for estimates of recombination rates. Due to the much smaller window size for GC% estimates, autocorrelation should decay quickly over short spatial scales. Such relatively fine-grain autocorrelation suggests that GC% estimates at the level of each gene should be relatively independent of one another. Therefore, we make no corrections for autocorrelation when including this variable in subsequent analyses.

Computational and statistical approaches

Data manipulation, autocorrelation and sliding window analyses were all performed using a mixture of MySQL database queries and custom Python scripts. All other analyses were performed in R v. 2.14.1 (R Development Core Team 2011). Specifically, we used regular multiple regression models for analyses focusing on the magnitude of the sex difference in recombination rate, the average recombination rate and the signed sex difference. Standard regression diagnostics were performed to assess model fit, and outlier influence and multicollinearity among the independent variables of the models was assessed by calculating variance inflation factors (Kutner et al. 2004). As the inclusion of variables in the model is based upon prior empirical work and theoretical predictions, all variables were kept in the linear models whether or not the model coefficient was significantly greater than zero. For analyses pertaining to the signed sex difference and the specific direction of bias for imprinted genes, a linear model was run for all variables except for the normal binary imprinting variable.

Residuals from this model were then used to assess any difference in the signed sex difference in recombination rate between paternally and maternally biased imprinted genes.

Results

Estimated imprinted genes

Using the more stringent two-fold expression difference threshold between maternally and paternally derived alleles, we estimated that 75 of the 12,625 genes investigated were imprinted. This number is roughly similar to previous estimates, as 68 of the 123 known imprinted genes derived from the Gene Imprint and Mousebook databases were also found in our overall sample of genes. These genes all received a numerical value of 1 if imprinted and 0 otherwise. Subsequent analyses break apart these imprinted genes into separate categories of maternally and paternally biased genes.

Main theoretical prediction

In order to test the main prediction that imprinted genes should be associated with larger recombinational sex differences, we constructed a linear model in which the spatially differenced absolute value of the sex difference in recombination was regressed onto our ten independent variables. These variables consisted of genes' imprinting status, their spatially differenced average recombination rate, and eight other genomic covariates described above. Separate linear models were constructed that utilized either the set of imprinted genes estimated here or the previously estimated set. Calculated variance inflation factors (VIF) indicated little to no

multicollinearity among these models' independent variables. As expected, each model is largely coincident with the other, with each model overall explaining a significant proportion of the variation in recombination rates (adjusted R^2 values of 0.248 and 0.247, respectively). In each analysis, four of the ten independent variables had regression coefficients that were significantly greater than zero (Tables 4.2, 4.3).

Table 4.2. Regression coefficients and associated test parameters for the analysis using new estimates of imprinted gene expression. Significant variables are shown in bold.

| Variable | Coefficient | Std. Error | t-statistic | p-value | VIF |
|------------------------------|----------------------|----------------------|--------------------|-----------------------|------------|
| (Intercept) | -0.003 | 0.044 | -0.079 | 0.937 | - |
| Imprinting | 0.146 | 0.038 | 3.821 | 0.0001 | 1.011 |
| Centromere distance | -0.043 | 0.010 | -4.181 | 2.92×10^{-5} | 1.044 |
| GC% | -0.0008 | 0.0009 | -0.947 | 0.344 | 1.469 |
| CpG distance | 1.5×10^{-8} | 2.3×10^{-8} | 0.699 | 0.485 | 1.058 |
| Compactness | -0.004 | 0.014 | -0.295 | 0.768 | 1.036 |
| LINE density | 0.158 | 0.094 | 1.672 | 0.094 | 1.527 |
| SINE density | 0.451 | 0.085 | 5.307 | 1.14×10^{-7} | 1.307 |
| LTR density | -0.170 | 0.092 | -1.848 | 0.065 | 1.046 |
| SREP density | -6.303 | 4.815 | -1.309 | 0.191 | 1.084 |
| Average recombination | 0.858 | 0.013 | 64.16 | $< 2 \times 10^{-16}$ | 1.018 |

Again as expected, the average local recombination rate had the largest effect in the model, as regions of low recombination are necessarily constrained to have small sex differences in their rates. However, the models also support previous theoretical predictions, with imprinted genes having a modest positive effect on the magnitude of sex differences in recombination.

Table 4.3. Regression coefficients and test parameters for the analysis using prior estimates of imprinted gene expression. Significant variables are shown in bold.

| Variable | Coefficient | Std. Error | t-statistic | p-value | VIF |
|------------------------------|-----------------------|-----------------------|-------------|-----------------------|-------|
| (Intercept) | -0.003 | 0.044 | -0.077 | 0.938 | - |
| Imprinting | 0.099 | 0.040 | 2.457 | 0.014 | 1.006 |
| Centromere distance | -0.043 | 0.010 | -4.200 | 2.69×10^{-5} | 1.052 |
| GC% | -0.0008 | 0.0009 | -0.952 | 0.3412 | 1.498 |
| CpG distance | 1.83×10^{-8} | 2.26×10^{-8} | 0.810 | 0.4181 | 1.055 |
| Compactness | -0.002 | 0.014 | -0.164 | 0.870 | 1.030 |
| LINE density | 0.166 | 0.094 | 1.757 | 0.079 | 1.550 |
| SINE density | 0.452 | 0.085 | 5.310 | 1.12×10^{-7} | 1.310 |
| LTR density | -0.170 | 0.092 | -1.840 | 0.066 | 1.037 |
| SREP density | -6.417 | 4.818 | -1.332 | 0.183 | 1.075 |
| Average recombination | 0.858 | 0.013 | 64.112 | $< 2 \times 10^{-16}$ | 1.027 |

The modest effects of significant variables in the models are most easily seen with standard partial regression coefficients, shown in Table 4.4 below. To facilitate comparison among variables in the model, each partial standard regression coefficient represents a standard deviation change in the magnitude of recombination sex differences for each increase of a standard deviation of that predictor variable (adjusting for all other variables in the model). As imprinting in the model is represented by a binary categorical variable, however, a slightly different standard regression coefficient was calculated instead to facilitate interpretation. In this case, the coefficient represents a standard deviation increase in the magnitude of recombinational sex differences between unimprinted (0) and imprinted genes (1).

Table 4.4. Standard partial regression coefficients for significant variables in both linear models. *Note the non-standard calculation of the coefficient for the imprinting variable as described above.

| Variable | Current imprinting estimates | Prior imprinting estimates |
|-----------------------|------------------------------|----------------------------|
| Average recombination | 0.500 | 0.500 |
| Imprinting* | 0.385 | 0.260 |
| SINE density | 0.047 | 0.047 |
| Centromere distance | -0.033 | -0.033 |

Imprinting and the direction of sex-biased recombination

The analyses above demonstrate that genomic imprinting has a non-negligible association with the magnitude of sex differences in recombination as predicted by theory. Given this result, we can then test a more subtle aspect of this association elaborated in a previous empirical study (Lercher and Hurst 2003). Specifically, recombination rates should be biased away from the sex whose allele gets expressed at an imprinted locus in the subsequent generation. In other words, paternally expressed genes should be characterized by relatively higher female recombination rates and vice versa. To address this prediction, we used the entire dataset to first regress the spatially differenced signed sex difference in recombination (female – male) on all the previous independent variables except for genes' imprinting status. We then calculated residuals from this linear model and used the subset corresponding to imprinted genes as the data for subsequent two sample tests. In terms of the theoretical predictions, imprinted genes that are maternally biased should therefore have a more negative residual value compared to paternally biased genes. Statistically, this corresponds to a one-tailed test of the null hypothesis of no association between the direction of imprinting and the sex difference in recombination.

For the set of imprinted genes estimated here, we performed two separate analyses. Visual inspection of the data and the results of Grubbs' outlier test (Grubbs 1969) highlighted a single outlier that could hold considerable leverage in our two sample tests (Figure 4.1). We therefore performed analyses with and without this single lowest value present in the sample. Due to the relatively small number of imprinted genes, we used nonparametric Wilcoxon rank sum tests to assess differences in residuals between maternally and paternally expressed genes.

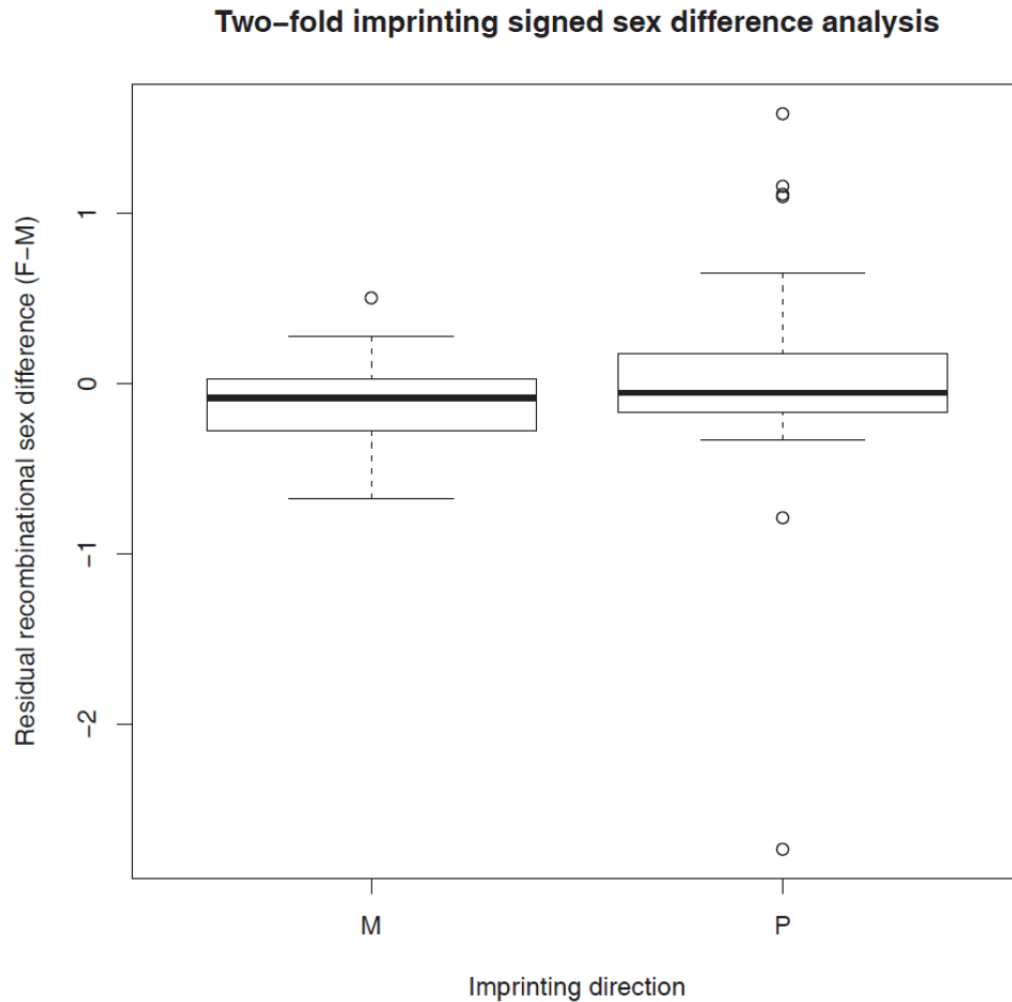


Figure 4.1. Box plots of regression residuals for maternally (M) and paternally (P) biased imprinted genes.

Without the outlier present, 28 maternally expressed and 45 paternally expressed genes were available for the test. Coincident with theoretical predictions, maternally expressed genes did indeed have a significantly smaller recombination residual (i.e. more male-biased recombination rates) than paternally expressed genes ($W = 471$, one-tailed $p = 0.036$). However, a similar analysis including the residual outlier was only marginally significant ($W = 499$, one-tailed $p = 0.054$).

Using previous estimates of imprinted, 38 maternally biased and 30 paternally biased genes were available for a similar analysis. For this set of genes, the Grubbs' test did not detect

any significant outliers (Figure 4.2), allowing us to perform only a single Wilcoxon test on the data. Again, maternally biased genes had significantly smaller residual values ($W = 406$, one-tailed $p = 0.021$) than paternally biased genes. Together, these results indicate that if such a difference is real, its magnitude may be relatively small. However, as was evident in the full regression models described above, substantial residual variation in recombination rates remains unexplained. Future accounting of presently latent variables will allow us to address the magnitude of imprinting's effects on recombination with far more precision.

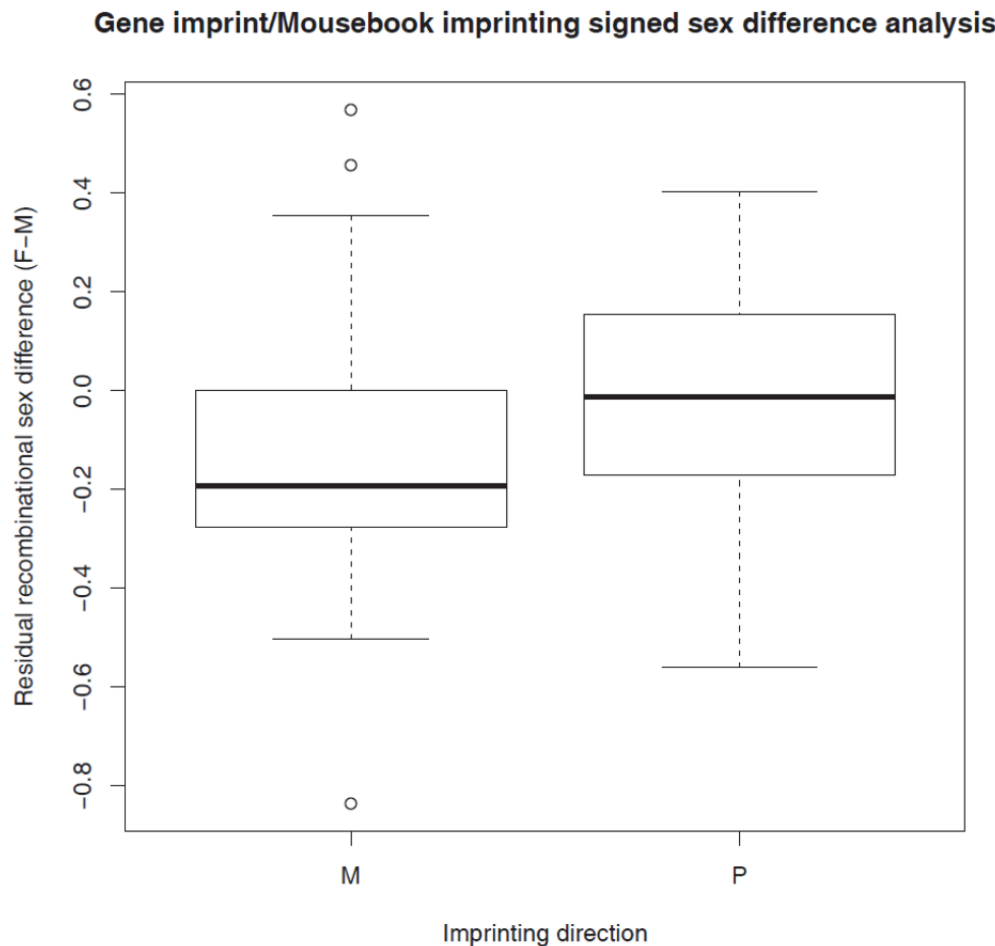


Figure 4.2. Box plots of regression residuals for maternally (M) and paternally (P) biased imprinted genes.

Discussion

Recombination is a complex cellular and evolutionary process that has presented numerous and deep conceptual problems to biologists for most of the history of genetics. Of these many problems, the existence of sex differences in recombination rates is one of the more unintuitive and empirically challenging to address. These difficulties have likely contributed to the paucity of attention paid to this phenomenon over time, relative to other aspects of recombination. However, population genetic modeling efforts in recent decades have made some progress in highlighting factors that would favor the evolution of these differences, a phenomenon also known as heterochiasmy (Lenormand 2003).

These modeling efforts have highlighted sex differences in haploid epistasis as the simplest and most general factor capable of favoring the evolution of heterochiasmy across many species. While such an explanation seems to offer little for taxa predominantly experiencing selection in the diploid phase of the life cycle, mammals represent a notable exception due to the phenomenon of genomic imprinting. Genes subject to imprinting are expressed solely from the maternally or paternally derived allele and thus can experience haploid selection during the diploid phase of the life cycle. Epistasis among imprinted genes, or potentially more commonly between imprinted and unimprinted genes (Joseph and Kirkpatrick 2004; Mott et al. 2014), can therefore fall within the purview of these theoretical predictions. Conveniently, such predictions can then be tested empirically provided estimates are available for both genomic imprinting and rates of sex-specific recombination.

Unfortunately, such joint estimates are rare in the existing literature. Indeed, prior empirical research on these predictions focused solely on humans (Lercher and Hurst 2003; Sandovici et al. 2006; Necsulea et al. 2009). Here, however, we have expanded on such previous

efforts, using newer data on genomic imprinting (Gregg et al. 2010) and higher resolution sex-specific recombination rates generated from experimental crosses of lab mice (Cox et al. 2009). While adjusting for a variety of genomic covariates, we indeed found a positive association between imprinting and local magnitudes of heterochiasmy as predicted by theory. Specifically, imprinting accounted for about a one-third standard deviation increase in the magnitude of sex differences in recombination (Table 4.4). Furthermore, associations between the directions of imprinting and sex-biased recombination rates also agreed with previous theoretical predictions. Paternally expressed genes exhibited significantly more female-biased residual recombination rates than maternally expressed genes, although this effect was modest (Figure 4.1, 4.2). But despite these positive associations and our inclusion of a set of variables known or suspected to influence recombination, most of the variation we observed in recombination differences between the sexes remained unexplained. What missing factors and aspects of the analysis might contribute to this unexplained variation in recombination?

Aside from possible underestimation of imprinting, theory itself provides us with some hint of the latent variables unaccounted for in our analyses. Although Lenormand's model draws attention to interacting imprinted genes experiencing quasi-haploid selection, other more complex forms of epistasis—including those normally operating in the diploid phase of the life cycle—contribute to heterochiasmy as well. Lenormand's interpretation of the model's results was purposely constructed from as broad a taxonomic perspective as possible, a constraint notably lacking in our analyses here. Mechanisms influencing heterochiasmy that cannot apply to hermaphrodites, for example, could nevertheless be more important to analyses here than one might expect from Lenormand's more general perspective. Taxonomic considerations aside, these other mechanisms may also simply occur less frequently within a genome or contribute

weaker selection pressures on recombination modifiers. Nonetheless, taking such forms of epistasis into account could in principle explain a small, though non-trivial proportion of our remaining residual variation. In practice, however, difficulties in measuring epistasis may render such explanatory variables effectively unmeasurable, bringing us back to where we started.

In stark contrast to epistasis, perhaps other eminently more measurable molecular aspects of the genome represent some of our latent explanatory variables. A cursory examination of the performance of the genomic variables we did include, however, suggests a degree of caution is warranted in gathering additional similar variables. Of the eight genomic covariates we fit in our regression models, only the density of SINE elements and the proportional distance to the centromere significantly explained a miniscule amount of the variation in recombination (Table 4.4). The remaining genomic variables therefore independently contributed effectively nothing to the analysis. Why might this be, given prior connections between these variables and recombination?

One issue could be the scale of such connections. Prior studies and our investigations here focused on broad-scale recombination rates (cM/Mb), whereas other studies have focused on hot-spots of recombination found at much smaller kilobase scales (Paigen and Petkov 2010; Smagulova et al. 2011). However, factors influencing small-scale aspects of recombination need not directly affect processes acting at larger scales. Features of DNA conducive to the production of double-stranded breaks, for instance, may have no similar influence on the shaping of chromatin along the synaptonemal complex or formation of recombination nodules. GC content, for example, was found to be significantly elevated within hot-spots, yet no correlation was observed at a broader level between chromosomal GC content and the density of hot-spots (Smagulova et al. 2011). Given the number of factors likely at play throughout the entire

recombination process, this scale-dependency is likely not restricted to GC content, complicating the effective use of other molecular variables in genomic analyses.

Methodological factors may also contribute to the persistent use of molecular variables that appear to have little explanatory power. First, such variables are relatively easy to measure, particularly in the well-annotated genomes of model organisms. Second, whole genome-scale analyses typically have massive sample sizes due to the large number of genes or other features considered. These samples provide most studies with enough power to detect biologically trivial associations between genomic variables with an absurd degree of statistical precision. Third, studies may fail to recognize the lack of independence among observations. Specifically, genes exist within the distinct spatial structures of their chromosomes. A molecular variable's observations may be determined in part by their neighbors or may vary anisotropically across the span of a chromosome. Genes influenced by such a variable would therefore no longer exhibit independence, biasing analyses predicated upon it. In the context of a linear regression, ignoring spatial dependencies can inflate the significance of explanatory variables by underestimating their standard errors (Martin 1974). Collectively, these factors have resulted in a set of genomic variables that may spuriously or weakly correlate with any other variable often enough that researchers may only continue measuring them for the sake of completeness. Furthermore, this problem is worsened by common implicit notions that biological phenomena ultimately are explicable solely in molecular terms (e.g. Barlow 1993). From this perspective, one should never truly dismiss a molecular variable—no matter how uninformative—as such variables are the only ones that matter.

Aside from sociological inertia in the genomics community, real aspects of the underlying biology also pose methodological challenges for such analyses. Despite my

unflattering appraisal of molecular variables above, aspects of genomic architecture logically must influence the process of crossing over. To the extent that we can or have measured such variables, however, we still lack a complete view of their relationship. One reason for this is that the regression approach we use implicitly treats causality as a unidirectional phenomenon. However, over evolutionary time, local patterns of crossing over may causally feed back on our explanatory genomic variables as well. Reciprocal causation obfuscates traditional distinctions between causes and effects, thereby complicating our efforts to derive meaning from the information such variables present (Haig 2014). Phylogenetically-explicit comparative approaches may represent one way out of this dilemma. Although relevant data for such methods is lacking, formally taking advantage of the unidirectional arrow of time might allow us to tease apart variation that is causal from that which has been caused.

A closer examination of the biology of recombination, particularly through the lens of theory, may also be fruitful in constructing a mechanistic basis for how sex differences might evolve. Specifically, in molecular terms, what is a modifier of recombination and how might it function? The focus here on broad-scale recombination rates suggests that answers may be found through an understanding of how chromatin is folded and arranged along the developing synaptonemal complex during prophase I.

Early in prophase I, chromosomes pair with their homologs and condense into visible chromatids with a particular shape and orientation. Specifically, chromosomes condense into a series of loops, each containing up to hundreds of kilobases of DNA, oriented perpendicularly to the developing axial elements of the future synaptonemal complex (Heng et al. 1996; Paigen and Petkov 2010). Double strand breaks can occur within inter-loop regions of DNA parallel to the axial element, with some fraction of these breaks later developing into mature crossovers

representing recombination between homologs (Paigen and Petkov 2010). Within a population, the lengths of the synaptonemal complex can vary and it has been shown this variation is causally linked to rates of recombination (Lynn et al. 2002). Synaptonemal complex lengths also differ between the sexes, with longer axes in females being associated with higher female recombination rates (Tease and Hultén 2004). Since the overall length of a chromosome is the same between the sexes and given the constant density of chromatin loops per physical length of synaptonemal complex (Kleckner et al. 2003), females therefore achieve higher recombination rates by constructing smaller, more numerous loops along their axial element (Figure 4.3).

A better molecular perspective on this differential loop construction may then narrow the set of features to consider in reconstructing how heterochiasmy may evolve. For example, *cis*-acting factors that vary in their ability to interact with lateral elements of the synaptonemal complex could be important. Such elements could destabilize loop formation in such a way as to shift regions of DNA into or out of the non-recombining loop chromatin. However, identifying such factors may be difficult depending on the range of such elements' effects. It is unclear whether or not loop modifications made at one position of the chromosome's axis remain localized or spread physically to neighboring loops. Downstream effects on neighboring loops would suggest that such *cis*-elements could be located at a substantial distance from the two loci whose recombination rate is being modified. Additionally, depending on the sensitivity of loop construction to chromosomal position (Heng et al. 1996), such *cis*-elements would potentially have pleiotropic consequences on recombination rates between a large set of other loci. Sex-specific *trans*-acting genes affecting chromatin states, however, might be more regionalized in their activity and thus could also represent plausible candidates for recombination modifiers. This notion is further supported by the fact that several meiotically-active genes are known to act

in a sex-specific manner (Hunt and Hassold 2002; Morelli and Cohen 2005). Finally, this potential diversity of *cis* and *trans*-acting factors would also represent a large mutational target, possibly explaining the observation that heterochiasmy evolves quickly (Lenormand and Dutheil 2005).

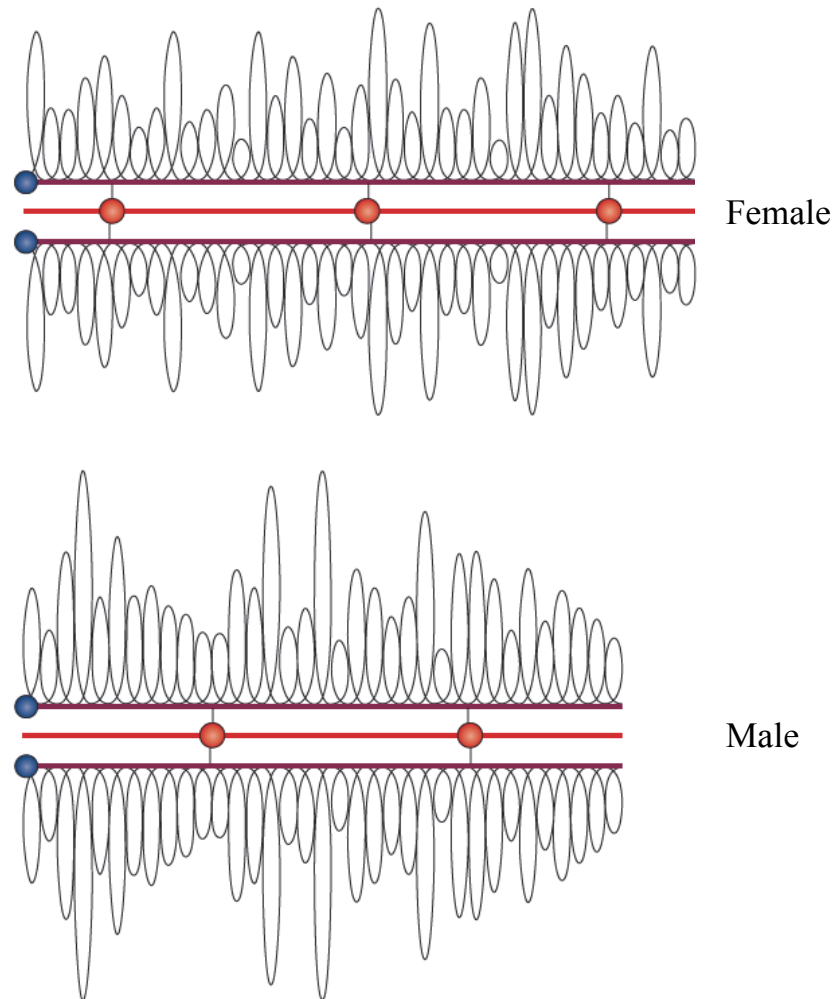


Figure 4.3. Representation of DNA loops (shown in black) and sex differences in the length of the synaptonemal complex (the three horizontal colored lines) for a given chromosome. Centromeres are represented by blue circles at the left edge of the image and crossover sites are represented by red circles along the central element of the synaptonemal complex. Modified from (Paigen and Petkov 2010).

Recombination as a genomic Ouroboros

Ultimately, a finer-scale understanding of the molecular basis of heterochiasmy can provide us with more concrete identities of the causal agents influencing recombination, shedding some light on its evolutionary lability as well. However, the arrow of causality need not terminate with the recombinational environment. Given this lability and specifically its sex-specific nature, what consequences might such recombination hold for the genome overall? As might be expected from a general understanding of the process, recombination mediates genetic notions of independence. However, the reciprocal action of heterochiasmy and genomic imprinting specifically can affect other, seemingly independent parts of the genome in a far more nuanced manner.

One such independent aspect of the genome is the set of genes experiencing sexually antagonistic selection. Beginning first with imprinted genes, epistasis between such genes may contribute to greater differences in the recombination rate between males and females (Lenormand 2003). Such changes to the local recombinational environment would be felt by other nearby genes. In particular, recent theoretical work has shown that sex differences in recombination can affect the ability of a sexually antagonistic allele to invade a population (Wyman and Wyman 2013). Other population genetic models have investigated the consequences of two independent linked genes respectively experiencing sexual and parental antagonism (Patten et al. 2013). One result of the model was that the presence of segregating sexually antagonistic variation at one locus can facilitate the maintenance of polymorphism at another nearby locus experiencing parental antagonism. One possible resolution of parental antagonism at a locus involves the evolution of imprinted gene expression. And finally,

provided that new imprinted gene interacts with other nearby genes, imprinted gene expression may spread to these epistatic partners (Wolf 2013). These epistatic interactions then bring us back full circle to the evolution of heterochiasmy.

In the scenario above, completely different aspects of a species' ecology may have independently caused these parentally and sexually antagonistic selection pressures. Mutations residing in offspring that favor paternal interests may be agnostic as to the sex of that offspring. Similarly, mutations benefitting females need not differentially affect matrilineal and patrilineal fitness. Yet depending on the specific history of mutation and selection, these separate pressures can become genetically correlated in such a way as to jointly influence future responses to other independent selective pressures. Specifically in our scenario, the local recombinational environment and its differences between the sexes, prior or otherwise, mediated this conflation of independent selective pressures. Prior imprinting and epistasis affected recombination, thereby affecting divergence between the sexes, further influencing parental divergence, resulting then in the evolution of new imprinted genes capable of modifying the recombinational landscape once more. In this way, recombination can be as much about building associations between disparate features of the genome as it is about breaking associations between alleles at smaller scales.

Conclusion

Based on the results shown here and in previous chapters, it seems clear that the organization and architecture of the genome is as much an agent of evolutionary change as it is a product of external evolutionary forces. However, this agent is less of a single, cohesive unit than we might often think, especially in the case of genomes sequentially experiencing distinct

sexual niches. The genome is instead a community of widely varying and incompletely independent evolutionary interests, a notion previously dubbed the ‘parliament of genes’ (Leigh 1971). This notion was previously invoked to explain the fairness of meiosis imposed by the many on the few who might actively distort the process of segregation, one of many forms of genetic conflict that genomes experience. Here, however, we stretch this metaphor a little farther in a different direction.

In our case, previously independent members of parliament (genes) may find their interests align as they experience sequential periods of economic boom and bust (male and female bodies), promoting coalition formation. Actions (phenotypes) of such coalitions made in response to the present state of the economy play heavily on the minds of voters tasked with judging their effectiveness (selection). Come election day, such voters determine the political future of the coalition *en masse*, with correlated consequences for each particular member’s subsequent term in office. Yet as voters’ memories stretch hardly at all into the past, the rise or fall of different coalitions over time substantially influences voters’ immediate perception of what might be normal or favorable. In this way, voters choose their favorites, yet it is the chosen that shape the context of future choice.

Despite the difficulties it may cause, such reciprocal causation should be more explicitly highlighted in evolutionary biology. Simple mental models of unidirectional causation have value in simplifying biological phenomena into logically clear and empirically tractable entities. However, we should not take this assumption’s utility as a sign that it accurately represents biological reality. At a minimum, circularity is inherent to biological systems as genes cause phenotypes and phenotypes cause genes to be represented in future generations (Haig 2014). Fostering a more explicitly ecological perspective on interactions within genomes or within the

individual would instead help us recognize the likely widespread extent of feedback relationships among biological phenomena. Understanding such phenomena would then require a broader consideration of seemingly disparate biological sub-disciplines, potentially sparing us from the tide of overspecialization readily apparent in biology today. In short, everyone in biology would benefit from recognizing and accepting the fundamental interconnectedness of all things.

Acknowledgments

I would like to thank Catherine Dulac and Chris Gregg for sharing this database of imprinted gene estimates and explaining its construction. I would also like to thank the reader for having successfully reached the end of this dissertation.

References

- Barlow, D. P. 1993. Methylation and imprinting: from host defense to gene regulation? *Science* 260: 309-310.
- Bonduriansky, R. and S. F. Chenoweth. 2009. Intralocus sexual conflict. *Trends in Ecology & Evolution* 24: 280-288.
- Brandvain, Y. and G. Coop. 2012. Scrambling eggs: meiotic drive and the evolution of female recombination rates. *Genetics* 190: 709-723.
- Burt, A., G. Bell and P. H. Harvey. 1991. Sex differences in recombination. *Journal of Evolutionary Biology* 4: 259-277.
- Chinwalla, A. T., L. L. Cook, K. D. Delehaunty, G. A. Fewell, L. A. Fulton, R. S. Fulton, T. A. Graves, L. W. Hillier, E. R. Mardis and J. D. McPherson. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420: 520-562.
- Cliff, A. D. and J. K. Ord. 1973. *Spatial autocorrelation*. Pion Limited, London, U.K.
- Cliff, A. D. and J. K. Ord. 1981. *Spatial processes: models & applications*. Pion, London, U.K.
- Corbett-Detig, R. B., J. Zhou, A. G. Clark, D. L. Hartl and J. F. Ayroles. 2013. Genetic incompatibilities are widespread within species. *Nature* 504: 135-137.
- Cosmides, L. M. and J. Tooby. 1981. Cytoplasmic inheritance and intragenomic conflict. *Journal of Theoretical Biology* 89: 83-129.
- Cox, A., C. L. Ackert-Bicknell, B. L. Dumont, Y. Ding, J. T. Bell, G. A. Brockmann, J. E. Wergedal, C. Bult, B. Paigen, J. Flint, S. W. Tsaih, G. A. Churchill and K. W. Broman. 2009. A new standard genetic map for the laboratory mouse. *Genetics* 182: 1335-1344.
- Dumont, B. L. and B. A. Payseur. 2011. Evolution of the genomic recombination rate in murid rodents. *Genetics* 187: 643-657.
- Dumont, B. L., M. A. White, B. Steffy, T. Wiltshire and B. A. Payseur. 2011. Extensive recombination rate variation in the house mouse species complex inferred from genetic linkage maps. *Genome Research* 21: 114-125.
- Eyre-Walker, A. and L. D. Hurst. 2001. The evolution of isochores. *Nature Reviews Genetics* 2: 549-555.
- Fujita, P. A., B. Rhead, A. S. Zweig, A. S. Hinrichs, D. Karolchik, M. S. Cline, M. Goldman, G. P. Barber, H. Clawson, A. Coelho, M. Diekhans, T. R. Dreszer, B. M. Giardine, R. A. Harte, J. Hillman-Jackson, F. Hsu, V. Kirkup, R. M. Kuhn, K. Learned, C. H. Li, L. R. Meyer, A. Pohl, B. J. Raney, K. R. Rosenbloom, K. E. Smith, D. Haussler and W. J. Kent. 2011. The UCSC Genome Browser database: update 2011. *Nucleic Acids Research* 39: D876-D882.

Geneimprint. Accessed June 2012. <http://www.geneimprint.org>.

Geraldes, A., P. Basset, B. Gibson, K. L. Smith, B. Harr, H.-T. YU, N. Bulatova, Y. Ziv and M. W. Nachman. 2008. Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. *Molecular Ecology* 17: 5349-5363.

Gregg, C., J. W. Zhang, B. Weissbourd, S. J. Luo, G. P. Schroth, D. Haig and C. Dulac. 2010. High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science* 329: 643-648.

Grubbs, F. E. 1969. Procedures for detecting outlying observations in samples. *Technometrics* 11: 1-21.

Guan, Y., C. L. Myers, R. Lu, I. R. Lemischka, C. J. Bult and O. G. Troyanskaya. 2008. A genomewide functional network for the laboratory mouse. *PLoS Computational Biology* 4: e1000165.

Haig, D. 1996. Placental hormones, genomic imprinting, and maternal—fetal communication. *Journal of Evolutionary Biology* 9: 357-380.

Haig, D. 2010. Games in tetrads: segregation, recombination, and meiotic drive. *The American Naturalist* 176: 404-413.

Haig, D. 2014. Fighting the good cause: meaning, purpose, difference, and choice. *Biology & Philosophy* 1-23.

Haldane, J. 1922. Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics* 12: 101-109.

Hartl, T. A., S. J. Sweeney, P. J. Knepler and G. Bosco. 2008. Condensin II resolves chromosomal associations to enable anaphase I segregation in *Drosophila* male meiosis. *PLoS Genetics* 4: e1000228.

Hassold, T. and P. Hunt. 2001. To err (meiotically) is human: the genesis of human aneuploidy. *Nature Reviews Genetics* 2: 280-291.

Heng, H., J. W. Chamberlain, X.-M. Shi, B. Spyropoulos, L.-C. Tsui and P. B. Moens. 1996. Regulation of meiotic chromatin loop size by chromosomal position. *Proceedings of the National Academy of Sciences* 93: 2795-2800.

Hunt, P. A. and T. J. Hassold. 2002. Sex matters in meiosis. *Science* 296: 2181-2183.

Iossifov, I., R. Rodriguez-Esteban, I. Mayzus, K. J. Millen and A. Rzhetsky. 2009. Looking at cerebellar malformations through text-mined interactomes of mice and humans. *PLoS Computational Biology* 5: e1000559.

- Jensen-Seaman, M. I., T. S. Furey, B. A. Payseur, Y. Lu, K. M. Roskin, C.-F. Chen, M. A. Thomas, D. Haussler and H. J. Jacob. 2004. Comparative recombination rates in the rat, mouse, and human genomes. *Genome Research* 14: 528-538.
- Joseph, S. B. and M. Kirkpatrick. 2004. Haploid selection in animals. *Trends in Ecology & Evolution* 19: 592-597.
- Kamme, F., R. Salunga, J. Yu, D.-T. Tran, J. Zhu, L. Luo, A. Bittner, H.-Q. Guo, N. Miller and J. Wan. 2003. Single-cell microarray analysis in hippocampus CA1: demonstration and validation of cellular heterogeneity. *The Journal of Neuroscience* 23: 3607-3615.
- Kleckner, N., A. Storlazzi and D. Zickler. 2003. Coordinate variation in meiotic pachytene SC length and total crossover/chiasma frequency under conditions of constant DNA length. *Trends in Genetics* 19: 623-628.
- Kutner, M. H., C. Nachtsheim and J. Neter. 2004. *Applied linear regression models*. McGraw-Hill/Irwin, Boston, MA.
- Lande, R. 1980. Sexual dimorphism, sexual selection, and adaptation in polygenic characters. *Evolution* 34: 292-305.
- Leigh, E. G. 1971. *Adaptation and diversity: natural history and the mathematics of evolution*. Freeman Cooper, San Francisco, CA.
- Lenormand, T. 2003. The evolution of sex dimorphism in recombination. *Genetics* 163: 811-822.
- Lenormand, T. and J. Dutheil. 2005. Recombination difference between sexes: a role for haploid selection. *PLoS Biology* 3: e63.
- Lercher, M. J. and L. D. Hurst. 2003. Imprinted chromosomal regions of the human genome have unusually high recombination rates. *Genetics* 165: 1629-1632.
- Li, X., H. Cai, J. Xu, S. Ying and Y. Zhang. 2010. A mouse protein interactome through combined literature mining with multiple sources of interaction evidence. *Amino Acids* 38: 1237-1252.
- Lynn, A., K. E. Koehler, L. Judis, E. R. Chan, J. P. Cherry, S. Schwartz, A. Seftel, P. A. Hunt and T. J. Hassold. 2002. Covariation of synaptonemal complex length and mammalian meiotic exchange rates. *Science* 296: 2222-2225.
- Mank, J. E. 2009. The evolution of heterochiasmy: the role of sexual selection and sperm competition in determining sex-specific recombination rates in eutherian mammals. *Genetics Research* 91: 355-363.
- Martin, R. L. 1974. On spatial dependence, bias and the use of first spatial differences in regression analysis. *Area* 6: 185-194.
- Mayr, E. 1961. Cause and effect in biology. *Science* 134: 1501-1506.

- McVean, G., L. D. Hurst and T. Moore. 1996. Imprinted genes have few and small introns. *Nature Genetics* 12: 234-237.
- Mochizuki, A., Y. Takeda and Y. Iwasa. 1996. The evolution of genomic imprinting. *Genetics* 144: 1283-1295.
- Moran, P. A. 1950. Notes on continuous stochastic phenomena. *Biometrika* 37: 17-23.
- Moran, P. A. P. 1948. Some theorems on time series: II the significance of the serial correlation coefficient. *Biometrika* 35: 255-260.
- Morelli, M. A. and P. E. Cohen. 2005. Not all germ cells are created equal: aspects of sexual dimorphism in mammalian meiosis. *Reproduction* 130: 761-781.
- Mott, R., W. Yuan, P. Kaisaki, X. Gan, J. Cleak, A. Edwards, A. Baud and J. Flint. 2014. The architecture of parent-of-origin effects in mice. *Cell* 156: 332-342.
- Mousebook. MRC Harwell. Accessed June 2012. <http://www.mousebook.org>.
- Necsulea, A., M. Sémon, L. Duret and L. D. Hurst. 2009. Monoallelic expression and tissue specificity are associated with high crossover rates. *Trends in Genetics* 25: 519-522.
- Novocraft. Novocraft Technologies. Accessed February 2014. <http://www.novocraft.com>.
- Paigen, K. and P. Petkov. 2010. Mammalian recombination hot spots: properties, control and evolution. *Nature Reviews Genetics* 11: 221-233.
- Pardo-Manuel de Villena, F. and C. Sapienza. 2001. Nonrandom segregation during meiosis: the unfairness of females. *Mammalian Genome* 12: 331-339.
- Parker, G. A., R. Baker and V. Smith. 1972. The origin and evolution of gamete dimorphism and the male-female phenomenon. *Journal of Theoretical Biology* 36: 529-553.
- Patten, M. M., F. Úbeda and D. Haig. 2013. Sexual and parental antagonism shape genomic architecture. *Proceedings of the Royal Society B: Biological Sciences* 280: 20131795.
- Petkov, P. M., K. W. Broman, J. P. Szatkiewicz and K. Paigen. 2007. Crossover interference underlies sex differences in recombination rates. *Trends in Genetics* 23: 539-542.
- Phillips, P. C. 2008. Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics* 9: 855-867.
- R Development Core Team. 2011. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at <http://www.R-project.org>.
- RepeatMasker Home Page. Institute for Systems Biology. Accessed July 2013. <http://www.repeatmasker.org>.

- Sandovici, I., S. Kassovska-Bratinova, J. E. Vaughan, R. Stewart, M. Leppert and C. Sapienza. 2006. Human imprinted chromosomal regions are historical hot-spots of recombination. *PLoS Genetics* 3: e101.
- Smagulova, F., I. V. Gregoret, K. Brick, P. Khil, R. D. Camerini-Otero and G. V. Petukhova. 2011. Genome-wide analysis reveals novel molecular features of mouse recombination hotspots. *Nature* 472: 375-378.
- Storey, J. D. and R. Tibshirani. 2003. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences* 100: 9440-9445.
- Tease, C. and M. A. Hultén. 2004. Inter-sex variation in synaptonemal complex lengths largely determine the different recombination rates in male and female germ cells. *Cytogenetic and Genome Research* 107: 208-215.
- Trivers, R. 1988. Sex differences in rates of recombination and sexual selection, pp. 270-286 in *The evolution of sex: an examination of current ideas*, edited by R. E. Michod and B. R. Levin. Sinauer Associates, Sunderland, MA.
- van Doorn, G. S. 2009. Intralocus sexual conflict. *Annals of the New York Academy of Sciences* 1168: 52-71.
- Wolf, J. B. 2013. Evolution of genomic imprinting as a coordinator of coadapted gene expression. *Proceedings of the National Academy of Sciences* 110: 5085-5090.
- Wyman, M. J. and M. C. Wyman. 2013. Sex-specific recombination rates and allele frequencies affect the invasion of sexually antagonistic variation on autosomes. *Journal of Evolutionary Biology* 26: 2428-2437.